

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:11:54 ; Search time 1073.96 Seconds

(without alignments)
1073.635 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264
Sequence: 1 ctgacttgagcaattccca.....gacatgctcatgatcttta 264

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_vl1:*
73: gb_vl2:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.2	19.4	7218	5	I66494
2	38.2	14.5	1552	48	ST095923
3	38.2	14.5	147098	39	AC011452
4	38.2	14.5	164313	39	AC010624
5	38.2	14.5	177590	50	AC021154
6	37.6	14.2	167390	10	AC007263
7	37.6	14.2	200542	71	CNS01RG3
8	37.4	14.2	163854	50	AC021590
9	37.2	14.1	79452	42	AC016536
10	37.2	14.1	107109	91	HS0J781B1
11	36.8	13.9	167079	67	AL160400
12	36.8	13.9	187999	54	AC026928

```

c 13 36.8 13.9 194963 51 AC022699
c 14 36.6 13.9 24721 91 HSL190H5
c 15 36.6 13.9 181517 68 AL353773
c 16 36.6 13.9 223263 67 AL158068
c 17 36.4 13.8 152592 40 AC012027
c 18 36.4 13.8 212382 39 AC010936
c 19 36.2 13.7 62109 38 AL359984
c 20 36 13.6 569 94 HUM075530
c 21 36 13.6 152370 66 AL138894
c 22 36 13.6 170132 71 CNS057DS
c 23 36 13.6 186841 40 AC011974
c 24 35.6 13.5 9070 7 AP148542
c 25 35.4 13.4 118591 43 AC020202
c 26 35.4 13.4 177987 31 AC008948
c 27 35.4 13.4 188942 32 AC026280
c 28 35.4 13.4 260580 32 AE003791
c 29 35.2 13.3 76042 7 AC006921
c 30 35.2 13.3 115725 53 AC025481
c 31 35.2 13.3 158546 51 AC022876
c 32 35.2 13.3 167312 70 AP001024
c 33 35.2 13.3 173081 59 AC068996
c 34 35.2 13.3 178817 68 AL354931
c 35 35.2 13.3 195944 70 AP002353
c 36 35.2 13.3 200038 43 AC018423
c 37 35 13.3 18788 7 AP131223
c 38 35 13.3 118357 49 HSL142C11
c 39 35 13.3 151102 69 AL391220
c 40 35 13.3 159747 70 AP001019
c 41 35 13.3 177055 70 AP001496
c 42 35 13.3 177629 58 AC060829
c 43 35 13.3 178676 59 AC026965
c 44 35 13.3 182945 54 AC069004
c 45 35 13.3 183005 67 AL161732

```

ALIGNMENTS

```

Db 1371 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1430
Qy 195 ttacgatacaatgctccca 214
Db 1431 YYYYYGTACCAAAATTCCTTA 1450

RESULT 2
STU95923 1552 bp mRNA PLN 11-MAY-1997
LOCUS Solanum tuberosum transaldolase (Pot1) mRNA, complete cds.
DEFINITION U95923
ACCESSION U95923
VERSION U95923.1 GI:2078349
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 1552)
AUTHORS Moehs,C.P., Allen,P.V., Friedman,M. and Belknap,W.R.
TITLE Cloning and expression of transaldolase from potato
JOURNAL Plant Mol. Biol. 32 (3), 447-452 (1996)
MEDLINE 97134933
REFERENCE 2 (bases 1 to 1552)
AUTHORS Moehs,C.P., Allen,P.V., Friedman,M. and Belknap,W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-1997) CIR, USDA/ARS, 800 Buchanan Street, Albany,
CA 94710, USA
FEATURES
source Location/Qualifiers
1..1552
/organism="Solanum tuberosum"
/strain="Lemhi Russet"
/db_xref="taxon:4113"
82..1398
/gene="Pot1"
82..1398
/gene="Pot1"
/gene="Pot1"
/EC_number="2.2.1.2"
/codon_start=1
/product="transaldolase"
/protein_id="AA054016.1"
/db_xref="GI:2078350"
/translating="MASCITKISNFCNENSTGYGSRSSSTPKFYVHLRTSPFNHLSGS
KLTLVINSIRASLCVRCQADNGSSVKKRTTLLHLYETGQSPWVDNLCPLVTDILP
LIDSGVGVTSNPAIFQKAISTSNVYNDQFRELVOAGKIDSVYWEVWELVLDQACKL
FETLYDKTDGDDGVSYVEVSPRLADDEGVYEAQKMLKKVRSNVIKIPATAPCIP
SIRFEVIGLISVNTLLFESLSRYAVIDAVLDGLDGLSGLSDLSRVTSVASFVSRVDT
LVDRKLEKIGTPPEALDRKRAANQALADYDIOKRSRPRMVAIVKGGKRRLLMA
STVKNRPVPTLVDPDIPDIPDIPVSTMPDQALDQFIDHGSVARTIDANLSEAGTISA
LEKIDVWVFVGSQLELGVDSFERSFSDLSLDSLQERANTLKLVLN"
BASE COUNT 439 a 315 c 343 g 455 t
ORIGIN

```

```

RESULT 1
LOCUS 166494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dornier,F., Schefflinger,F. and Falkner,F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14-23-SEP-1997;
FEATURES
source Location/Qualifiers
1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

```

Query Match 19.4%; Score 51.2; DB 5; Length 7218;
 Best Local Similarity 5.5%; Pred. No. 0.00036;
 Matches 11; Conservative 128; Mismatches 61; Indels 0; Gaps 0;

```

Qy 15 ttccaatgctccgttccaactccaagccaactcactgcttcttaccgaa 74
Db 1251 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1310
Qy 75 ggcgtcgcccccagatctcgcttcctcactccaactcttcacatgctttaa 134
Db 1311 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1370
Qy 135 cacaataactatcgctccctccgacatcactcactgctccctcttaaacctc 194

```

```

Qy 181 ctcttaaacctccctacgataatgctcccaacagatggagatgctcgcaa 240
Db 239 CTATTAGGCGCATCTTGTGTCAATAGCTCCCAAGCTGATGGAATGGAAGCTGATGA 298
Qy 241 agagagac 247
Db 299 AGGAGAC 305

RESULT 3
AC011452 147098 bp DNA HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 19 clone CTC-326K19, WORKING DRAFT
DEFINITION SEQUENCE, 3 ordered pieces.

```

```

ACCESSION      AC011452
VERSION        AC011452.5
KEYWORDS       HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 147098)
JOURNAL        Sequencing of Human Genome Institute.
REFERENCE      2 (bases 1 to 147098)
AUTHORS        Unpublished
TITLE          DOE Joint Genome Institute.
JOURNAL        DOE Joint Genome Institute.
COMMENT        Direct Submission
               Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
               On May 6, 2000 this sequence version replaced gi:7690094.
               -----Genome Center
               Center: Joint Genome Institute
               Center code: JGI
               Web site: http://www.jgi.doe.gov
               -----
               Project Information
               Center Project Name: 310622, BC275645
               Center clone name: CIT-HSPC_326K19
               -----
               Summary Statistics
               Consensus quality: 145829 bases at least Q40
               Consensus quality: 146451 bases at least Q30
               Consensus quality: 146700 bases at least Q20
               Estimated insert size: 147530; agarose-fp estimation
               Estimated insert size: 147048; sum-of-ctdigs estimation
               Quality coverage: 6.64 in Q20 bases; agarose-fp estimation
               Quality coverage: 6.66 in Q20 bases; sum-of-ctdigs estimation
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 3 contigs. Gaps between the contigs
               * are represented as runs of 'N'. The order of the pieces
               * is believed to be correct as given, however the sizes
               * of the gaps between them are based on estimates that have
               * provided by the submittor.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.
               * 1 22742: contig of 22742 bp in length
               * 22743 22842: gap of unknown length
               * 22843 80776: contig of 57934 bp in length
               * 80777 80876: gap of unknown length
               * 80877 147098: contig of 66222 bp in length.
               Location/Qualifiers
               1. 147098
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="19"
               /clone_lib="CalTech human BAC library C"
               /clone_1lb="37789 c 37738 g 36142 t 200 others
BASE COUNT      35229 a 37789 c 37738 g 36142 t 200 others
ORIGIN
Query Match      14.5%; Score 38.2; DB 39; Length 147098;
Best Local Similarity 61.6%; Pred. No. 1.4;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 97 gccctccacactcaaacctcttcacatgcgttcaacacaaactacatcgctt 156
    || || || || || || || || || || || || || || || || || ||
Db 6591 GCCCTCTTCCTCAGAGCGCTTCTCCACACCCGCTGTATATAAAGCAAGCAATCTACCGCTGG 7050
    || || || || || || || || || || || || || || || || || ||
QY 157 cccttcgattcatcactcgtcctccctcttaaacctc 195
    || || || || || || || || || || || || || || || || || ||
Db 7051 CCTGCCATTCCTTCCTCCATTAATCCTTAATACGCGCT 7089

```

```

AC010624/C
LOCUS
DEFINITION
AC010624 164313 bp DNA HTG 21-JUN-2000
Homo sapiens chromosome 19 clone CTD-2126E3, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC010624
AC010624.5 GI:8576043
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 164313)
TITLE
DOE Joint Genome Institute.
JOURNAL
Sequencing of Human Chromosome 19
REFERENCE
AUTHORS
2 (bases 1 to 164313)
TITLE
DOE Joint Genome Institute.
JOURNAL
Direct Submission
COMMENT
Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 21, 2000 this sequence version replaced gi:7690069.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 676032, BC641056
Center clone name: CITB-HL_2126E3
-----
Summary Statistics
Consensus quality: 157887 bases at least Q40
Consensus quality: 161345 bases at least Q30
Consensus quality: 162314 bases at least Q20
Estimated insert size: 212000; agarose-ftp estimation
Estimated insert size: 163713; sum-of-contigs estimation
Quality coverage: 11.38 in Q20 bases; agarose-ftp estimation
Quality coverage: 14.74 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1041: contig of 1041 bp in length
1042 1141: gap of unknown length
1142 5695: contig of 4554 bp in length
5696 5795: gap of unknown length
5796 14933: contig of 9138 bp in length
14934 15034: gap of unknown length
15034 25674: contig of 10641 bp in length
25675 25774: gap of unknown length
25775 44213: contig of 18439 bp in length
44214 44314: gap of unknown length
44314 61176: contig of 16863 bp in length
61177 61276: gap of unknown length
61277 164313: contig of 103037 bp in length.
*
* Location/Qualifiers
1. 164313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2126E3"
/clone_11b="CalTech human BAC library D"
BASE COUNT
43310 a 39759 c 39049 g 41593 t 602 others
ORIGIN
Query Match 14.5%: Score 38.2; DB 39; Length 164313;
Best Local Similarity 61.6%: Pred. No. 1.4;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```



```
misc_feature
misc_feature
```

BASE COUNT
ORIGIN

Db 29909

SOURCE ORGANISM

REFERENCE
AUTHORS

COMMENT

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3000
Center clone name: 9_B.13

* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 855 954: gap of 100 bp in length
* 955 1851: contig of 897 bp in length
* 1852 1951: gap of 100 bp in length
* 1952 2858: contig of 907 bp in length
* 2859 2958: gap of 100 bp in length
* 2959 3854: contig of 896 bp in length
* 3855 3954: gap of 100 bp in length
* 3955 4608: contig of 854 bp in length
* 4809 4908: gap of 100 bp in length
* 4909 5774: contig of 866 bp in length
* 5775 5874: gap of 100 bp in length
* 5875 6750: contig of 876 bp in length
* 6751 6850: gap of 100 bp in length
* 6851 7734: contig of 884 bp in length
* 7735 7834: gap of 100 bp in length
* 7835 8745: contig of 911 bp in length
* 8746 8845: gap of 100 bp in length
* 8846 9749: contig of 904 bp in length
* 9750 9849: gap of 100 bp in length
* 9850 10559: contig of 710 bp in length
* 10560 10659: gap of 100 bp in length
* 10660 11556: contig of 897 bp in length
* 11557 11656: gap of 100 bp in length
* 11657 12548: contig of 892 bp in length
* 12549 12648: gap of 100 bp in length
* 12649 13500: contig of 852 bp in length
* 13501 13600: gap of 100 bp in length
* 13601 14485: contig of 885 bp in length
* 14486 14585: gap of 100 bp in length
* 14586 15450: contig of 865 bp in length
* 15451 15550: gap of 100 bp in length
* 15551 16437: contig of 887 bp in length
* 16438 16537: gap of 100 bp in length
* 16538 17410: contig of 873 bp in length
* 17411 17510: gap of 100 bp in length
* 17511 18409: contig of 899 bp in length
* 18410 18509: gap of 100 bp in length
* 18510 19392: contig of 883 bp in length
* 19393 19492: gap of 100 bp in length
* 19493 20375: contig of 883 bp in length
* 20376 20475: gap of 100 bp in length
* 20476 21364: contig of 889 bp in length
* 21365 21464: gap of 100 bp in length
* 21465 22332: contig of 868 bp in length
* 22333 22432: gap of 100 bp in length
* 22433 23293: contig of 861 bp in length
* 23294 23393: gap of 100 bp in length
* 23394 24284: contig of 891 bp in length
* 24285 24384: gap of 100 bp in length
* 24385 25268: contig of 884 bp in length
* 25269 25368: gap of 100 bp in length
* 25369 26262: contig of 894 bp in length
* 26263 26362: gap of 100 bp in length
* 26363 27262: contig of 900 bp in length
* 27263 27362: gap of 100 bp in length
* 27363 28266: contig of 904 bp in length
* 28267 28366: gap of 100 bp in length

* 28367 29229: contig of 863 bp in length
* 29230 29329: gap of 100 bp in length
* 29330 30239: contig of 910 bp in length
* 30240 30339: gap of 100 bp in length
* 30340 31213: contig of 874 bp in length
* 31214 31313: gap of 100 bp in length
* 31314 32207: contig of 894 bp in length
* 32208 32307: gap of 100 bp in length
* 32308 33194: contig of 887 bp in length
* 33195 33294: gap of 100 bp in length
* 33295 34169: contig of 875 bp in length
* 34170 34269: gap of 100 bp in length
* 34270 35132: contig of 863 bp in length
* 35133 35232: gap of 100 bp in length
* 35233 36110: contig of 878 bp in length
* 36111 36210: gap of 100 bp in length
* 36211 37099: contig of 889 bp in length
* 37100 37199: gap of 100 bp in length
* 37200 38065: contig of 866 bp in length
* 38066 38165: gap of 100 bp in length
* 38166 39050: contig of 885 bp in length
* 39051 39150: gap of 100 bp in length
* 39151 40062: contig of 912 bp in length
* 40063 40162: gap of 100 bp in length
* 40163 41044: contig of 882 bp in length
* 41045 41144: gap of 100 bp in length
* 41145 42060: contig of 916 bp in length
* 42061 42160: gap of 100 bp in length
* 42161 43051: contig of 891 bp in length
* 43052 43151: gap of 100 bp in length
* 43152 44041: contig of 890 bp in length
* 44042 44141: gap of 100 bp in length
* 44142 44996: contig of 855 bp in length
* 44997 45096: gap of 100 bp in length
* 45097 45958: contig of 863 bp in length
* 45960 46059: gap of 100 bp in length
* 46060 46945: contig of 886 bp in length
* 46946 47045: gap of 100 bp in length
* 47046 47946: contig of 901 bp in length
* 47947 48046: gap of 100 bp in length
* 48047 48918: contig of 873 bp in length
* 48920 49019: gap of 100 bp in length
* 49020 49920: contig of 901 bp in length
* 49921 50020: gap of 100 bp in length
* 50021 50898: contig of 878 bp in length
* 50899 50998: gap of 100 bp in length
* 50999 51891: contig of 893 bp in length
* 51892 51991: gap of 100 bp in length
* 51992 52893: contig of 902 bp in length
* 52894 52993: gap of 100 bp in length
* 52994 53889: contig of 896 bp in length
* 53890 53989: gap of 100 bp in length
* 53990 54867: contig of 878 bp in length
* 54868 54967: gap of 100 bp in length
* 54968 55811: contig of 844 bp in length
* 55812 55911: gap of 100 bp in length
* 55912 56786: contig of 875 bp in length
* 56787 56886: gap of 100 bp in length
* 56887 57787: contig of 901 bp in length
* 57788 57887: gap of 100 bp in length
* 57888 58783: contig of 886 bp in length
* 58784 58883: gap of 100 bp in length
* 58884 59770: contig of 887 bp in length
* 59771 59870: gap of 100 bp in length
* 59871 60782: contig of 912 bp in length
* 60783 60882: gap of 100 bp in length
* 60883 61771: contig of 889 bp in length
* 61772 61871: gap of 100 bp in length
* 61872 62779: contig of 908 bp in length
* 62780 62879: gap of 100 bp in length
* 62880 63759: contig of 880 bp in length
* 63760 63859: gap of 100 bp in length
* 63860 64764: contig of 905 bp in length

[illegible][illegible]

```
misc_feature
1. .6987
/note="assembly_fragment:00219
clone_end:SP6
vector_side:left"
7088. .25926
/note="assembly_fragment:00622
fragment_chain:1"
26027. .27279
/note="assembly_fragment:00819
fragment_chain:1"
27360. .44222
/note="assembly_fragment:00646
fragment_chain:1"
44333. .69220
/note="assembly_fragment:00797
fragment_chain:1"
69321. .81427
/note="assembly_fragment:01531
fragment_chain:1"
81528. .105379
/note="assembly_fragment:01300
fragment_chain:1"
105480. .129641
/note="assembly_fragment:00126
fragment_chain:1"
129742. .139328
/note="assembly_fragment:01328
```

Eumakriota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
 1 (bases 1 to 187999)
 Birtner, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 9, clone RP11-18A3
 Unpublished
 2 (bases 1 to 187999)
 Birtner, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
 Boguski, L., Bouckgeater, B., Brown, A., Burkett, G.,
 Campiano, A., Castle, A., Chappel, Y., Colangelo, M., Collins, S.,
 Collins, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domini, M., Doyle, M., Ferreira, P., Fitzhugh, L.,
 Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Haeos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lechoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margolis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Melchior, J., Meunier, L., Milhova, T., Miranda, C., Mlepta, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olliver, T. M., Oliver, J., Peterson, K., Pierré, N.,
 Pisanti, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

TITLE
JOURNAL
COMMENT

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teafaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 22, 2000 this sequence version replaced g1:7328793.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 18_A3
Center clone name: 18_A3

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167725 bases at least Q40
Consensus quality: 178025 bases at least Q30
Consensus quality: 182047 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 184399; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1228 1327: contig of 1227 bp in length
1328 1327: gap of 100 bp
1328 2416: contig of 1089 bp in length
2417 2516: gap of 100 bp
2517 4057: contig of 1541 bp in length
4058 4157: gap of 100 bp
4158 5171: contig of 1014 bp in length
5172 5271: gap of 100 bp
5272 6380: contig of 1109 bp in length
6381 6480: gap of 100 bp
6481 7659: contig of 1179 bp in length
7660 7759: gap of 100 bp
7760 9214: contig of 1455 bp in length
9215 9314: gap of 100 bp
9315 10932: contig of 1618 bp in length
10933 11032: gap of 100 bp
11033 12035: contig of 1003 bp in length
12036 12135: gap of 100 bp
12136 13215: contig of 1080 bp in length
13216 13315: gap of 100 bp
13316 14645: contig of 1330 bp in length
14646 14745: gap of 100 bp
14746 15925: contig of 1180 bp in length
15926 16025: gap of 100 bp
16026 17158: contig of 1133 bp in length
17159 17258: gap of 100 bp
17259 19445: contig of 2187 bp in length
19446 19545: gap of 100 bp
19546 21242: contig of 1697 bp in length
21243 21342: gap of 100 bp
21343 24035: contig of 2693 bp in length
24036 24135: gap of 100 bp
24136 25142: contig of 1007 bp in length
25143 25242: gap of 100 bp

FEATURES

source

25243 27117: contig of 1875 bp in length
27118 27217: gap of 100 bp
27218 28912: contig of 1655 bp in length
28913 29012: gap of 100 bp
29013 31508: contig of 2456 bp in length
31509 31608: gap of 100 bp
31609 32697: contig of 1089 bp in length
32698 32797: gap of 100 bp
32798 37275: contig of 4478 bp in length
37276 37375: gap of 100 bp
37376 40980: contig of 3605 bp in length
40981 41080: gap of 100 bp
41081 45215: contig of 4135 bp in length
45216 45315: gap of 100 bp
45316 48710: contig of 3395 bp in length
48711 48810: gap of 100 bp
48811 52850: contig of 4040 bp in length
52851 52950: gap of 100 bp
52951 57072: contig of 4122 bp in length
57073 57172: gap of 100 bp
57173 62417: contig of 5245 bp in length
62418 62517: gap of 100 bp
62518 71007: contig of 8490 bp in length
71008 71107: gap of 100 bp
71108 79805: contig of 8698 bp in length
79806 79905: gap of 100 bp
79906 81160: contig of 1255 bp in length
81161 81260: gap of 100 bp
81261 92353: contig of 11093 bp in length
92354 92453: gap of 100 bp
92454 101125: contig of 8672 bp in length
101126 101225: gap of 100 bp
101226 117073: contig of 15848 bp in length
117074 117173: gap of 100 bp
117174 136020: contig of 18847 bp in length
136021 136120: gap of 100 bp
136121 161828: contig of 25708 bp in length
161829 161928: gap of 100 bp
161929 187999: contig of 26071 bp in length.
Location/Qualifiers
1. 187999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9"
/clone="RP11-18A3"
/clone_id="RPC1-11 Human Male BAC"
1. 1227
/note="assembly_fragment"
1328. 2416
/note="assembly_fragment"
2517. 4057
/note="assembly_fragment"
4158. 5171
/note="assembly_fragment"
5272. 6380
/note="assembly_fragment"
6481. 7659
/note="assembly_fragment"
7760. 9214
/note="assembly_fragment"
9315. 10932
/note="assembly_fragment"
11033. 12035
/note="assembly_fragment"
12136. 13215
/note="assembly_fragment"
13316. 14645
/note="assembly_fragment"
14746. 15925
/note="assembly_fragment"
16026. 17158
/note="assembly_fragment"

Query Match	13.9%;	Score 36.8;	DB 54;	Length 187999;
Best Local Similarity	53.5%;	Pred. No. 3.3;		
Matches 77; Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0;

Qy 185 taaacctcttaacgatacaatg 208
 | | | | | | | |
 Db 119439 ccaggcccatgacctctcacactg 119462

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 194963)	Bitren, B., Linton, L., Nusbaum, C. and Lander, E.	Mus musculus, clone RP23-10718	Unpublished	
2 (bases 1 to 194963)	Bitren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.			

1 (bases 1 to 194963)
Mammalia, Eutheria, Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-10718
Unpublished
2 (bases 1 to 194963)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barra, N., Beckery, R., Beda, F.,
Boguslavsky, L., Boucknight, B., Brown, A., Burkett, G., Castle, A.,
Choepl, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domingo, M., Doyle, M., Feneclor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hacos, B., Heaford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Katata, A., Klein, J.,
Macdonald, P., Lebeck, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K.,
Mcneeters, R., Meldrum, I., Meneses, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, D.,
Pierre, N., Pisan, C., Pollara, Y., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stenger-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Wu, A., Wu, X., Wyman, D., Ye, W. J.,

45620 50053: contig of 4434 bp in length

NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 50054 50153: gap of 100 bp
* 50154 55501: contig of 5348 bp in length
* 55502 55601: gap of 100 bp
* 55602 60490: contig of 4889 bp in length
* 60491 60590: gap of 100 bp
* 60591 66444: contig of 5854 bp in length
* 66445 66544: gap of 100 bp
* 66545 71111: contig of 4567 bp in length
* 71112 71211: gap of 100 bp
* 71212 77867: contig of 6656 bp in length
* 77868 77967: gap of 100 bp
* 77968 82747: contig of 4780 bp in length
* 82748 82847: gap of 100 bp
* 82848 88691: contig of 5844 bp in length
* 88692 88791: gap of 100 bp
* 88792 95730: contig of 6939 bp in length
* 95731 95830: gap of 100 bp
* 95831 103037: contig of 7207 bp in length
* 103038 103137: gap of 100 bp
* 103138 111460: contig of 8323 bp in length
* 111461 111560: gap of 100 bp
* 111561 119148: contig of 7588 bp in length
* 119149 119248: gap of 100 bp
* 119249 128097: contig of 8849 bp in length
* 128098 128197: gap of 100 bp
* 128198 137747: contig of 9550 bp in length
* 137748 137847: gap of 100 bp
* 137848 149328: contig of 11481 bp in length
* 149329 149428: gap of 100 bp
* 149429 160664: contig of 11236 bp in length
* 160665 160764: gap of 100 bp
* 160765 170402: contig of 9638 bp in length
* 170403 170502: gap of 100 bp
* 170503 182201: contig of 11699 bp in length
* 182202 182301: gap of 100 bp
* 182302 194963: contig of 12662 bp in length.

```

FEATURES

source

```

misc_feature 1..1103
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature 1204..2643
             /note="assembly_fragment"
misc_feature 2744..3758
             /note="assembly_fragment"
misc_feature 3859..4988
             /note="assembly_fragment"
misc_feature 5089..6259
             /note="assembly_fragment"
misc_feature 6360..8126
             /note="assembly_fragment"
misc_feature 8227..9555
             /note="assembly_fragment"
misc_feature 9656..11586
             /note="assembly_fragment"
misc_feature 11687..13441
             /note="assembly_fragment"
misc_feature 13542..14986
             /note="assembly_fragment"
misc_feature 15087..18988
             /note="assembly_fragment"
misc_feature 19089..21553
             /note="assembly_fragment"
misc_feature 21654..24665
             /note="assembly_fragment"
misc_feature 24766..29058
             /note="assembly_fragment"
misc_feature 29159..31541
             /note="assembly_fragment"
misc_feature 31642..35036
             /note="assembly_fragment"

```

```

misc_feature 35137..38278
             /note="assembly_fragment"
misc_feature 38379..41421
             /note="assembly_fragment"
misc_feature 41522..45519
             /note="assembly_fragment"
misc_feature 45620..50053
             /note="assembly_fragment"
misc_feature 50154..55501
             /note="assembly_fragment"
misc_feature 55602..60490
             /note="assembly_fragment"
misc_feature 60591..66444
             /note="assembly_fragment"
misc_feature 66545..71111
             /note="assembly_fragment"
misc_feature 71212..77867
             /note="assembly_fragment"
misc_feature 77968..82747
             /note="assembly_fragment"
misc_feature 82848..88691
             /note="assembly_fragment"
misc_feature 88792..95730
             /note="assembly_fragment"
misc_feature 95731..95830
             /note="assembly_fragment"
misc_feature 95831..103037
             /note="assembly_fragment"
misc_feature 103038..103137
             /note="assembly_fragment"
misc_feature 103138..111460
             /note="assembly_fragment"
misc_feature 111461..111560
             /note="assembly_fragment"
misc_feature 111561..119148
             /note="assembly_fragment"
misc_feature 119149..119248
             /note="assembly_fragment"
misc_feature 119249..128097
             /note="assembly_fragment"
misc_feature 128098..128197
             /note="assembly_fragment"
misc_feature 128198..137747
             /note="assembly_fragment"
misc_feature 137748..137847
             /note="assembly_fragment"
misc_feature 137848..149328
             /note="assembly_fragment"
misc_feature 149329..149428
             /note="assembly_fragment"
misc_feature 149429..160664
             /note="assembly_fragment"
misc_feature 160665..160764
             /note="assembly_fragment"
misc_feature 160765..170402
             /note="assembly_fragment"
misc_feature 170403..170502
             /note="assembly_fragment"
misc_feature 170503..182201
             /note="assembly_fragment"
misc_feature 182202..182301
             /note="assembly_fragment"
misc_feature 182302..194963
             /note="assembly_fragment"

```

Query Match 13.9%; Score 36.8; DB 51; Length 194963;
 Best Local Similarity 56.7%; Pred. No. 3.3;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

Oy 28 ccgttccaaagcttcaacgcaatccacttcttccatcagaagcttcgtccccc 87
    || || || || || || || || || || || || || || || || || || || ||
Db 84605 CAGCTTCCAGATGTCAAAGAACAGCAGTGCCTCCCTTGGATGCTTGCTTGCATTC 84546
    || || || || || || || || || || || || || || || || || || || ||
Oy 88 gagattccggtctctctacaccttcttccatcgcgtttatcaacaaactaac 147
    || || || || || || || || || || || || || || || || || || || ||
Db 84545 CCTCTCACCAAGTCACCTTCGCAAGTCTCCCGTCGCTCGACGACGCGGCAACC 84486
    || || || || || || || || || || || || || || || || || || || ||

```

```

RESULT 14
HSL190H5 24721 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from cosmid L190H5, Huntington's Disease Region,
DEFINITION Chromosome 4p16.3.
ACCESSION Z68163.1 GI:1122279
VERSION 268163
KEYWORDS 4p16.3; CPG island; repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 24721)
AUTHORS Mungall, A.
TITLE Direct submission
JOURNAL Submitted (05-DEC-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, England. E-mail enquiries: humpub@sanger.ac.uk
COMMENT On Dec 14, 1995 this sequence version replaced gi:1107691.
IMPORTANT:
This sequence is not the entire insert of clone L190H5. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
The true left end of clone L190H5 is at 1 in this sequence. The
true left end of clone L129H7 is at 24672 in this sequence. L190H5
is from cosmid library LAD4NCO1 constructed at the Human Genome
Center, Los Alamos National Laboratory, NM 87345 under the auspices
of the U.S. Department of Energy. The library was constructed using
flow-sorted human chromosome 4 from a Hamster-Human hybrid cell
line (UV20H121-27) containing human chromosomes 4, 8 and 21.
VECTOR: scot1
L190H5 is contained in a clone contig spanning
2mb which is described in Baxendale et al, Nature Genetics 4 (1993
) 181-186. See also Myers et al, Cytogenet Cell Genet. 66 (1994
) 218-230.
Location/Qualifiers

```

FEATURES

Location/Qualifiers

source	1. .24721
repeat_region	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="4" /map="4p16.3" /clone="LA04NC01" /cell_line="UV20HL21-27" /clone="LA04NC01-19045" 321. .448
repeat_region	/note="MIR2 repeat: matches 134. .4 of consensus" 515. .590
repeat_region	/note="2 copies of 38 mer 88 & conserved" 942. .1055
repeat_region	/note="L1mel1 repeat: matches 281. .388 of consensus" 1703. .1863
repeat_region	/partial /note="AluJo repeat: matches 138. .298 of consensus" 2251. .2521
repeat_region	/partial /note="AluSq repeat: matches 270. .1 of consensus" 4301. .4375
misc_feature	/note="15 copies of 5 mer 84 & conserved" 4394. .5404
repeat_region	/note="putative CpG island" 5151. .5430
repeat_region	/note="AluY repeat: matches 4. .294 of consensus" 5650. .5954
repeat_region	/note="AluSx repeat: matches 300. .1 of consensus" 6385. .6657
misc_feature	/note="AluJo repeat: matches 288. .5 of consensus" 8582. .9726
repeat_region	/note="match: VNTR X60679; Polymorphic by comparison with X60679" 12702. .12819
repeat_region	/partial /note="AluJo repeat: matches 118. .1 of consensus" 13898. .14167
repeat_region	/note="L1MB8 repeat: matches 914. .635 of consensus" 14541. .14626
repeat_region	/note="L1MB4 repeat: matches 931. .840 of consensus" 14605. .14642
repeat_region	/note="MADE1 repeat: matches 38. .1 of consensus" 14644. .14768
repeat_region	/note="L1MB6 repeat: matches 659. .535 of consensus" 14808. .15360
repeat_region	/note="L1 repeat: matches 4385. .4947 of consensus" 15391. .15483
repeat_region	/note="MLT2FA repeat: matches 1. .93 of consensus" 15487. .16397
repeat_region	/note="L1MA7 repeat: matches 1036. .88 of consensus" 16397. .16576
repeat_region	/note="L1 repeat: matches 3795. .3979 of consensus" 16630. .16875
repeat_region	/note="L1 repeat: matches 3977. .4223 of consensus" 16888. .17170
repeat_region	/partial /note="AluSg repeat: matches 283. .1 of consensus" 17175. .18225
repeat_region	/note="L1 repeat: matches 4207. .5291 of consensus" 18247. .18541
repeat_region	/note="AluJo repeat: matches 1. .302 of consensus" 19778. .20074
repeat_region	/note="MLT2B repeat: matches 77. .406 of consensus" 20188. .20586
repeat_region	/note="L1PA2 repeat: matches 493. .891 of consensus" 20639. .20839
repeat_region	/note="L1 repeat: matches 5179. .5389 of consensus" 20696. .21209
repeat_region	/note="L1ME1 repeat: matches 5. .518 of consensus" 21232. .21543
repeat_region	/note="L1MB5 repeat: matches 562. .922 of consensus" 24260. .24349
repeat_region	/note="10 copies of 9 mer 90 & conserved"

[illegible]

RESULT	15
LOCUS	AL353773/c
DEFINITION	AL353773 181517 bp DNA HTG , 24-MAY-2000 Homo sapiens chromosome 9 clone RP11-574M5, *** SEQUENCING IN PROGRESS ***, 12 unordered pieces.
ACCESSION	AL353773
VERSION	AL353773.1 GI:9213204
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 181517)
Plumb B.
Direct Submission
Submitted (22-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbquery@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Jul 15, 2000 this sequence version replaced q1:765336.

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgberys@sanger.ac.uk
Project information
Center project name: ba574m5
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; l08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 174373 bases at least Q40
Consensus quality: 177734 bases at least Q30
Consensus quality: 179410 bases at least Q20
Insert size: 180417; sum-of-contigs
Insert size: 194280; agarose-fp
Quality coverage: 3.87x in Q20 bases; sum-of-contigs
Quality coverage: 3.59x in Q20 bases; agarose-fp

```

* NOTE: This is a "working draft" sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 7042: contig of 7042 bp in length
* 7043 7142: gap of 100 bp

```
* 7143 15239: contig of 8097 bp in length
* 15240 15339: gap of 100 bp
* 15340 40312: contig of 24973 bp in length
* 40313 40412: gap of 100 bp
* 40413 41496: contig of 1084 bp in length
* 41497 41596: gap of 100 bp
* 41597 45493: contig of 3897 bp in length
* 45494 45593: gap of 100 bp
* 45594 81525: contig of 35932 bp in length
* 81526 81625: gap of 100 bp
* 81626 105239: contig of 23614 bp in length
* 105240 105339: gap of 100 bp
* 105340 132375: contig of 27036 bp in length
* 132376 132475: gap of 100 bp
* 132476 142831: contig of 10356 bp in length
* 142832 142931: gap of 100 bp
* 142932 144114: contig of 1183 bp in length
* 144115 144214: gap of 100 bp
* 144215 156872: contig of 12658 bp in length
* 156873 156972: gap of 100 bp
* 156973 181517: contig of 24545 bp in length.
```

FEATURES

SOURCE

```
1. 181517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-574M5"
/clone_1fb="RPC1-11.2"
1. 7042
/note="assembly_fragment:00172"
clone_end:"7"
vector_side:"left"
7143..15239
/note="assembly_fragment:00673"
fragment_chain:1"
15340..40312
/note="assembly_fragment:00607"
fragment_chain:1"
40413..41496
/note="assembly_fragment:00878"
fragment_chain:1"
41597..45493
/note="assembly_fragment:01120"
fragment_chain:2"
45594..81525
/note="assembly_fragment:00551"
fragment_chain:2"
81626..105239
/note="assembly_fragment:00551"
fragment_chain:2"
105340..132375
/note="assembly_fragment:00357"
fragment_chain:3"
132476..142831
/note="assembly_fragment:00439"
fragment_chain:3"
142932..144114
/note="assembly_fragment:00325"
144215..156872
/note="assembly_fragment:00621"
156973..181517
/note="assembly_fragment:00649"
BASE COUNT 54727 a 36050 c 35860 g 53777 t 1103 others
ORIGIN
```

```
Query Match 13.9%; Score 36.6; DB 68; Length 181517;
Best Local Similarity 53.1%; Pred. No. 3.8;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
```

```
QY 52 atccactgtcttcattcagagcgcttcgccccgagatctcgtcttcacttca 111
111 11 11 111 111 111 111 111 11 11 11 11 11 11 11 11
Db 129772 ATCTCTGACCTCGTATCCGCCGCTTGTGCTCCAGAGTGTGGATTACAGGTGTG 129713
```

```
QY 112 aactcttccatcgctttaaatacaaaactaactatccgcttcctccgattccattc 171
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 129712 AGCCACCGTGGCTGGCTGTGATCTCTTTAACCACTCACTTCCACCAACCAAT 129653
QY 172 actcgcttcctttaaactccttac 198
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 129652 ACCCGCTCCACCAACCAACCAACCAAC 129626
```

Search completed: November 11, 2000, 18:38:17
Job time: 1583 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:15:50 ; Search time 74.01 Seconds

(without alignments)

1340.021 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264

Sequence: 1 ctgactgacgaatccca.....gacatgctcatgacttta 264

Scoring table:

IDENTITY_NIC
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID56/gcgdata/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32.4	12.3	857	20	X57401
2	31.4	11.9	1612	19	V44595
3	31.4	11.9	4002	21	Z60603
4	31.2	11.8	5357	19	T94101
5	31.2	11.8	5357	17	T18551
6	31.2	11.8	5357	19	T94108
7	30.8	11.7	36901	20	Z23892
8	30.8	11.7	38886	20	Z23897
9	29.8	11.3	1798	18	T51756
10	29.6	11.2	1036	20	Z19421
11	29.6	11.2	1036	20	Z19209
12	29.6	11.2	1664976	19	V21209

13	29.4	11.1	3935	21	Z89458
14	29.2	11.1	12793	19	V44251
15	29	11.0	1005	20	Z15507
16	29	11.0	5752	20	Z30052
17	28.8	10.9	525	18	T67902
18	28.8	10.9	1444	21	Z93353
19	28.8	10.9	2575	16	O89558
20	28.8	10.9	2575	16	O90331
21	28.6	10.8	433	20	X52010
22	28.6	10.8	549	21	A15937
23	28.6	10.8	696	20	X97697
24	28.6	10.8	966	21	Z36244
25	28.6	10.8	997	21	A15947
26	28.6	10.8	2388	20	V84554
27	28.4	10.7	910715	20	X20248
28	28.4	10.7	1239	21	Z44075
29	28.2	10.7	1361	21	A02089
30	28.2	10.7	1451	21	Z56720
31	28	10.6	1593	21	A02504
32	28	10.6	2253	10	N92071
33	27.8	10.5	614	21	Z80128
34	27.8	10.5	1496	21	Z52486
35	27.6	10.5	758	19	V44215
36	27.6	10.5	1235	20	Z16147
37	27.6	10.5	1432	16	O90211
38	27.6	10.5	2035	19	V30292
39	27.6	10.5	3131	18	T91708
40	27.6	10.5	3562	19	V60839
41	27.6	10.5	3562	21	Z61197
42	27.6	10.5	3647	19	V60842
43	27.6	10.5	3647	21	Z61200
44	27.6	10.5	18443	20	V83942
45	27.4	10.4	420	21	Z80329

ALIGNMENTS

RESULT 1
X57401/c
ID X57401 standard; DNA; 857 BP.
XX
AC X57401:
XX
XX 24-JUL-1999 (first entry)
XX
DE Rat U3 gene trap derived nucleic acid 14A13E-RE.
XX
XX Gene trap; rat; cellular gene; viral infection; cell survival; cancer;
XX
XX Tumour progression; suppression; identification; viral growth;
XX
XX Tumour suppressor; prevention; screening; therapeutic agent; ss.
XX
XX Rattus norvegicus.
XX
XX W09919481-A2.
XX
XX 22-APR-1999.
XX
XX 08-OCT-1998: 98WO-US21276.
XX
XX 10-OCT-1997: 97US-0062021.
XX
XX (UYVA-) UNIV VANDERBILT.
XX
XX Dubois RN, Organ EL, Rubin DH;
XX
XX WPI; 1999-326546/27.
XX
XX Nucleic acid encoding tumor suppressors and products required for
XX
XX viral infection
XX
XX Claim 1; Page 58-59; 94pp; English.
XX


```

RESULT 5
T18551
ID T18551 standard; DNA: 53577 BP.
XX
AC T18551;
XX
DT 06-MAY-1997 (first entry)
XX
DE Human polycystic kidney disease normal PKD1 gene.
XX
KM Adult onset polycystic kidney disease; APKD; autosomal dominant;
KM mutant; transversion; transition; deletion; insertion; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 4379..5272
FT /tag= a
FT /note= "specifically claimed region of intronless
FT cDNA identified by exon trapping"
FT old_sequence replace(50652..50653, cg)
FT /tag= b
FT /note= "changes Val codon to Leu codon"
FT old_sequence replace(50796..50797, cg)
FT /tag= c
FT /note= "replaces Val codon by Leu codon"
FT old_sequence insert(51827..51828, cc)
FT /tag= d
FT /note= "insertion, results in frameshift"
XX
XX W09612033-A1.
XX
XX 25-APR-1996.
XX
XX 11-OCT-1995; 95WO-US13357.
XX
XX 31-JAN-1995; 95US-0381520.
XX
XX 12-OCT-1994; 94US-0323443.
XX
XX (TGIG-) IG LAB INC.
XX (OYUO ) UNIV JOHNS HOPKINS.
XX
XX Burn TC, Connors TD, Dackowski W, Germino G, Klinger KW;
XX Landes GM, Qian F;
XX WPI; 1996-222017/22.
XX
XX Isolated human polycystic kidney disease gene and its mutants
XX useful for treatment of polycystic kidney disease and screening for
XX carriers
XX
XX Claim 1; Fig 1; 65pp; English.
XX
XX The present sequence is that of the normal human PKD1 gene from
XX chromosome 16. Mutations in this gene (e.g. translocations,
XX transversions, deletions and/or insertions) are associated with
XX adult-onset polycystic kidney disease (APKD). The PKD1 locus is
XX GC-rich (62.4%). Comparison of this sequence with a previously
XX reported partial cDNA sequence revealed differences at three
XX locations (see features table). The most significant difference is
XX the presence of two additional cytosine residues on the plus-strand
XX at position 4566 of the previously reported sequence. The insertion
XX results in a frame-shift in the predicted protein coding sequence,
XX leading to replacement of 92 C-terminal amino acids with a novel
XX 12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.
XX There is a region consisting of 17 tandem copies of a perfect 27 bp
XX repeat and two large CT-rich regions.
XX
XX Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T; 0 other;

```

Query Match 11.8%; Score 31.2; DB 17; Length 53577;
 Best Local Similarity 47.4%; Pred. No. 3.3;

```

Matches 93; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 18 ccaatgcttcgcgttccaagctctcaacgcgaatcacttgcttcttcaagaagcg 77
  ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35270 cccatgccccccctccctccctccctccctccctccctccctccctccctcc 35329
QY 78 ctgcgccccagatctcgccttcacatccaacatttccatcgctttaaoc 137
  ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35330 atccccctctctctccctctctctccctctctctccctccctctctccct 35389
QY 138 aaactaactatcgcgttccctccgcatcactgcgtctcctttaaactcctta 197
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35390 ctctccctccctcccatccctccctccctccctccctccctccctccct 35449
QY 198 cggatcaatgctccc 213
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35450 cctccccaactcctcc 35465

```

```

RESULT 6
T94108
ID T94108 standard; DNA: 53577 BP.
XX
XX T94108;
XX
XX 01-JUN-1998 (first entry)
XX
XX Human PKD1 locus between chromosomal markers ATP6C and D16s84.
XX
XX Human; polycystic kidney disease 1; PKD1; treatment;
XX autosomal dominant polycystic kidney disease; APKD; ss.
XX
XX Homo sapiens.
XX
XX W09744457-A1.
XX
XX 27-NOV-1997.
XX
XX 22-MAY-1997; 97WO-US08799.
XX
XX 03-JUN-1996; 96US-0658136.
XX
XX 24-MAY-1996; 96US-0655360.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Burn T, Connors T, Dackowski W, Germino G, Klinger K;
XX Qian F;
XX WPI; 1998-018511/02.
XX
XX Human polycystic kidney disease gene, PKD1 - useful to treat and
XX diagnose human autosomal or adult onset polycystic kidney disease
XX
XX Example 5; Pages 60-89; 257pp; English.
XX
XX The present sequence is the human polycystic kidney disease 1
XX (PKD1) locus between chromosomal markers ATP6C and D16s84.
XX The PKD1 gene or polypeptide may be used to treat autosomal
XX dominant polycystic kidney disease (APKD), and identify carriers
XX of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
XX (Ab) that distinguish between normal and mutant PKD1 sequences can
XX also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
XX perform subcellular and histochemical localisation studies, and to
XX block the function of PKD1. Ab are also useful in rational drug
XX design studies to identify and test inhibitors of PKD1. Sense and
XX antisense sequences derived from the PKD1 gene may be used for
XX detection and therapy.
XX
XX Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T; 0 other;

```

Query Match 11.8%; Score 31.2; DB 19; Length 53577;
 Best Local Similarity 47.4%; Pred. No. 3.3;

	Matches	93:	Conservative	0:	Mismatches	103:	Indels	0:	Gaps	0:
QY	18	ccaatggtcttcggtttcccaagctctcaacgcgaacacacatctgcttccttatacgaagcg	77							
Db	35269	cccatggccccctccctccctccctcccatcccccctccctccctccctccctccctcc	35328							
QY	78	ctgcgcccccgagattctcgtctctcaacttcaacctcttcacatgcttttatcac	137							
Db	35329	atccccctctctctccctccctctccctccctccctccctccctccctccctccctcc	35388							
QY	138	aaaactaaactatctcgcgtccctccgataccatctcgaatactcgtccctctttaaaccctta	197							
Db	35389	ctcctcccccctccctcccatcccccctccctcccatcccccctccctccctccctcc	35448							
QY	198	cggatccaatgctcc	213							
Db	35449	ctctcccaactctctc	35464							

XX	RESULT 7
XX	223892/c
XX	ID 223892 standard; DNA; 36901 BP.
XX	223892;
XX	25-JAN-2000 (first entry)
XX	
XX	Murine LOBO genomic DNA fragment 2.
XX	
KW	LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW	diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW	spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
OS	Mus musculus.
XX	
PN	W09950284-A2.
XX	
PD	07-OCT-1999.
XX	
PF	26-MAR-1999; 99WO-EP02055.
XX	
PR	27-MAR-1998; 98DE-1013799.
XX	
PA	(ROSE/) ROSENTHAL A.
XX	
PI	Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX	WPI; 1999-601320/51.
DR	
XX	
PT	Nucleic acids encoding proteins which influence bone development,
XX	useful for treating and studying bone disorders -
PS	
XX	Example 3; Page 97-117; 391pp; German.
XX	
CC	This invention describes novel nucleic acids (I; designated LOBO (long
CC	bones)) encoding proteins influencing bone development in mammals. The
CC	products of the invention reduce and/or inactivate bone extension (i.e.
CC	development), with exception of the skull and have osteopathic activity.
CC	The nucleic acid molecules, proteins and antibodies can be used in
CC	diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC	and nucleic acid molecules, etc. are useful for production of transgenic
CC	animals, especially a transgenic mouse for the study of diseases
CC	associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC	achondroplasia. This sequence encodes the murine LOBO protein described
CC	in the method of the invention.
XX	
XX	Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;

[illegible]

XX	RESULT	8
XX	ID	Z23897/c
XX		Z23897 standard; DNA; 3886 BP.
XX	AC	Z23897;
XX	DT	25-JAN-2000 (first entry)
XX	DE	Murine LOBO homologue genomic DNA fragment 3.
XX	KW	LOBO; long bones; bone development; bone extension; skull; osteopathic;
XX	KM	diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
XX	KW	spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX	OS	Mus musculus.
PN		WO950284-A2.
PD		07-OCT-1999.
PF		26-MAR-1999; 99WO-EP02055.
PR		27-MAR-1998; 98DE-1013799.
PA		(ROSE/) ROSENTHAL A.
PI		Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
DR		WPI; 1999-601320/51.
XX		Nucleic acids encoding proteins which influence bone development,
XX		useful for treating and studying bone disorders -
PS		Example 3; Page 189-211; 391pp; German.
CC		This invention describes novel nucleic acids (if designated LOBO (long
CC		bones)) encoding proteins influencing bone development in mammals. The
CC		proteins of the invention reduce and/or inactivate bone extension (i.e.
CC		development), with exception of the skull and have osteopathic activity.
CC		The nucleic acid molecules, proteins and antibodies can be used in
CC		diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC		and nucleic acid molecules, etc. are useful for production of transgenic
CC		animals, especially a transgenic mouse for the study of diseases
CC		associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC		achondroplasia. This sequence encodes the murine LOBO protein described
CC		in the method of the invention.
XX		Sequence 3886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other:

<p>...nucleotide sequence of the murine LOBO protein described in the method of the invention.</p> <p>X X</p> <p>Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;</p>		<p>Query Match 11.7%; Score 30.8; DB 20; Length 36901; Best local Similarity 48.8%; Pred. No. 3.9; Matches 83; Conservative 0; Mismatches 87; Indels 0; Gaps 0;</p>	
QY	16 tcccaatggcttcggttcacaagctcaacgcaatccactgtcttcattacgaag 75	QY	16 tcccaatggcttcggttcacaagctcaacgcaatccactgtcttcattacgaag 75
Db	10309 TCCTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 10250	Db	10309 TCCTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 10250
QY	76 cgcttgcgcccgagattctgcgttcctcaactcaaaccttcctccatcgttttaac 135	QY	76 cgcttgcgcccgagattctgcgttcctcaactcaaaccttcctccatcgttttaac 135

XX Claim 4; Page 249; 299pp; English.

CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. 219249 to 219460 and Y39083 to Y39225
CC are used in the exemplification of the present invention.

XX Sequence 1036 BP; 153 A; 303 C; 177 G; 183 T; 220 other;

Query Match 11.2%; Score 29.6; DB 20; Length 1036;
Best Local Similarity 29.7%; Pred. No. 2.5;
Matches 62; Conservative 37; Mismatches 110; Indels 0; Gaps 0;

QY 5 ctggcagcaattcccaatggtcgttcctccaagctcctaagcgaatccactgtctc 64
DB 727 ctcksaamcwnccnctckkaacnccccckykgsmtncaccaatntwcmckcnsktntm 786
QY 65 ctatcagaagcgctcgccccgagattcgtctcctcactcctaacttctccat 124
DB 787 ctcccaaytnccnccnccrctccckstckstcamwtataaaccwcyawynkknccwm 846
QY 125 cgctttaaatacaaaactaactatccgtctccctcgcatactcactgtccctct 184
DB 847 awlamgacwctcnyncnccnckntktlamwccckmccckcswtwcykscsccmctct 906
QY 185 taaacctcttaagcatcaatgctccc 213
DB 907 macyccccckktynkwmcccttcccc 935

RESULT 11
ID 219209
XX 219209 standard; CDNA; 1036 BP.

AC 219209;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen cDNA encoding 5' Erdn-8.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity; ss.

OS Mycobacterium tuberculosis.

PN W09942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

PA (CORI-) CORIYA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
DR WPI: 1999-527416/44.

PT New polypeptide comprising antigenic portions of M. tuberculosis

XX Claim 4; Page 294-295; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

XX Sequence 1036 BP; 153 A; 303 C; 177 G; 183 T; 220 other;

Query Match 11.2%; Score 29.6; DB 20; Length 1036;
Best Local Similarity 29.7%; Pred. No. 2.5;
Matches 62; Conservative 37; Mismatches 110; Indels 0; Gaps 0;

QY 5 ctggcagcaattcccaatggtcgttcctccaagctcctaagcgaatccactgtctc 64
DB 727 ctcksaamcwnccnctckkaacnccccckykgsmtncaccaatntwcmckcnsktntm 786
QY 65 ctatcagaagcgctcgccccgagattcgtctcctcactcctaacttctccat 124
DB 787 ctcccaaytnccnccnccrctccckstckstcamwtataaaccwcyawynkknccwm 846
QY 125 cgctttaaatacaaaactaactatccgtctccctcgcatactcactgtccctct 184
DB 847 awlamgacwctcnyncnccnckntktlamwccckmccckcswtwcykscsccmctct 906
QY 185 taaacctcttaagcatcaatgctccc 213
DB 907 macyccccckktynkwmcccttcccc 935

RESULT 12
ID 219209/c
XX 219209 standard; DNA; 1664976 BP.

AC 219209;

DT 10-NOV-1998 (first entry)

DE Methanococcus jannaschii circular chromosome.

KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW genome; autotrophic; extrachromosomal element; identification; ds.

OS Methanococcus jannaschii.

PN W09807830-A2.

PD 26-FEB-1998.

PF 22-AUG-1997; 97WO-US14900.

PR 22-AUG-1996; 96US-0024428.

PA (GENO-) INST GENOMIC RES.

PA (UNIT) UNIV ILLINOIS FOUND.

PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;

DR WPI: 1998-169145/15.

PT Complete genome sequence of methano-genic archaeon, Methanococcus
PT jannaschii - useful in identification of M. jannaschii genome
PT fragment

XX Claim 13; Page 152-585; 614pp; English.

CC The present sequence represents the complete 1.66-megabase pair genome
CC sequence of the Methanococcus jannaschii circular chromosome. The
CC present invention describes M. jannaschii open reading frames from the

CC genome sequence. The invention also describes a computer based system
 CC for identifying fragments of the M. jannaschii genome that are
 CC homologous to target nucleotide sequences, comprising: (a) data storage
 CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550
 CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
 CC least 99.9% identical to it; (b) search means for comparing a target
 CC sequence to the nucleotide sequence of the data storage means to
 CC identify a homologous sequence, and (c) retrieval means for obtaining
 CC the homologous sequence. The method, which is based on whole genome
 CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
 CC of which consists of 3 physically distinct elements, a large circular
 CC chromosome (the 1664976 bp sequence given in V21209), a large circular
 CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a
 CC small circular extra-chromosomal element (the 16550 bp sequence given in
 CC V21211), can be used in the identification of M. jannaschii genome
 CC fragment.
 CC
 SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Query Match 11.2%; Score 29.6; DB 19; Length 1664976;
 Best Local Similarity 48.3%; Pred. No. 37;
 Matches 83; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

OY 25 ctccggttccaaagctcctaagccaaatccactgcttcctatccagaagcgtccgc 84

DB 1452938 CTTTAAATTCTGAAATCATCTTATCTGGAACCTCAACAAAAGCCTTTAAAA 1452879

OY 85 ccgagagattcgtctccaccctcaaacctctccatccgtcttcaaaccaacta 144

DB 1452878 CCTCAACAATTCCTCTTAGTATCTTACCTCGAATAGAGCTTTAAAGGCTCTCTTA 1452819

OY 145 acctatcgcctccctccgcatccatccatcgcctccctttaaaccctct 196

DB 1452818 AATGCTATCTCTCAACTATCATATCATATCTCTTAAATCTCCTT 1452767

RESULT 13

ID 289458 standard; DNA; 3935 BP.

XX 289458;

DT 16-JUN-2000 (first entry)

XX Murine trans-synaptic tracer protein encoding DNA #1.

XX Trans-synaptic tracer protein; murine; transgenic animal; nerve cell;

KM screening; nervous system; trans-synaptic tracer protein; ds.

XX Mus musculus.

OS Location/Qualifiers

XX Key 1369..2770

XX CDS /tag= a

XX exon /note= "trans-synaptic tracer protein"

XX /tag= b

XX /number= 1

XX /tag= c

XX /number= 1

XX /tag= d

XX /number= 2

XX /tag= e

XX /number= 2

XX /tag= f

XX /number= 3

PN WO200010382-A1.

XX 02-MAR-2000.

XX 18-AUG-1999; 99WO-JP04439.

XX 19-AUG-1998; 98JP-0232817.

XX (CHUS) CHUGAI SEIYAKU KK.

XX Yoshihara Y;

XX WPI: 2000-237506/20.

XX P-PsDB; Y51907.

XX Transgenic animal used for screening compounds for possible nervous

XX system effects has specific nerve cell types with transformed, with a

XX Claim 3; Page 25-28; 33pp; Japanese.

XX This invention describes a novel transgenic animal (I) in which
 CC specific nerve cell types have been transformed with a gene for a
 CC trans-synaptic tracer protein. An independent claim is also included
 CC for screening compounds with neurological activity using (I). The
 CC products of the invention can also be used for screening of compounds
 CC of the nervous system. This sequence encodes a murine trans-synaptic
 CC tracer protein which is used in the method described in the invention.
 CC
 SQ Sequence 3935 BP; 978 A; 988 C; 1024 G; 945 T; 0 other;

Query Match 11.1%; Score 29.4; DB 21; Length 3935;
 Best Local Similarity 56.8%; Pred. No. 4.8;
 Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 44 aacgcaaatcactgtctctctatcagaagcgttcgccccgagatctcgtctct 103

DB 3256 aattcacaatcctctgtgatacatcctcgtgagtgatgaaatcgaattacagctctct 3315

OY 104 cacttcaaacctctccatccatccgtctttaaaccaca 138

DB 3316 ctccgtctcctctctctatcccatcagccaca 3350

RESULT 14

ID V44251/C standard; DNA; 12793 BP.

XX V44251;

DT 19-JUL-1999 (first entry)

XX Lettuce resistance gene RG25.

XX Resistance gene; RG25; lettuce; disease resistance;

KM pest resistance; virus; fungus; protozoan; bacterium; nematode;

KW crop protection; transgenic plant; fingerprinting; ss.

XX Lactuca sativa.

OS Location/Qualifiers

XX Key 190..12577

XX CDS /tag= a

XX exon /note= "contains introns"

XX /tag= b

XX /number= 1

XX /tag= c

XX /number= 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:14:35 ; Search time 75.1 Seconds
(without alignments)
531.661 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264
Sequence: 1 ctgactgcagcaattccca.....gacatgctcatgattctta 264

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/5C.COMB.seq:*
4: /cgn2_6/pdata/1/lna/5D.COMB.seq:*
5: /cgn2_6/pdata/1/lna/6.COMB.seq:*
6: /cgn2_6/pdata/1/lna/CTUS.COMB.seq:*
7: /cgn2_6/pdata/1/lna/Dackfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.2	19.4	7218	1	US-08-232-463-14
2	31.4	11.9	4002	5	US-09-356-952-9
3	31.2	11.8	53526	5	US-08-658-136-2
4	31.2	11.8	53577	5	US-08-658-136-1
5	28.8	10.9	2575	1	US-08-135-511-33
6	28.8	10.9	2575	1	US-08-483-852-10
7	28.8	10.9	2575	1	US-08-361-458-5
8	28.8	10.9	2575	1	US-08-477-953-10
9	28.8	10.9	2575	2	US-08-187-453-33
10	28.8	10.9	2575	3	US-08-477-952-10
11	27.6	10.5	3131	5	US-09-035-648-23
12	27	10.2	933	7	5340934-12
13	27	10.2	2444	3	US-08-821-355A-2
14	27	10.2	2444	4	US-09-003-687A-2
15	26.8	10.2	1255	2	US-08-518-878B-38
16	26.8	10.2	1255	2	US-08-294-522B-38
17	26.8	10.2	1255	2	US-08-470-868A-38
18	26.8	10.2	1545	1	US-08-400-275-17
19	26.8	10.2	1596	5	US-08-807-861A-38
20	26.8	10.2	1596	5	US-09-210-681-38
21	26.6	10.1	564	7	US-08-946-719A-38
22	26.6	10.1	564	7	5175695-2
23	26.6	10.1	1931	1	US-08-530-492-2
24	26.4	10.0	325	3	US-08-332-766A-11
25	26.4	10.0	848	5	US-09-009-913-338
26	26.4	10.0	2246	6	PCT-US91-08525-23

27	26.4	10.0	2246	6	PCT-US91-08525-27	Sequence 27, Appl
28	26.4	10.0	2246	6	PCT-US91-08525-33	Sequence 33, Appl
29	26.4	10.0	2246	6	PCT-US93-04384-3	Sequence 3, Appl
30	26.4	10.0	2246	6	PCT-US93-04384-5	Sequence 5, Appl
31	26.4	10.0	3716	6	PCT-US91-07046-1	Sequence 1, Appl
32	26.4	10.0	4359	3	US-08-566-398-23	Sequence 23, Appl
33	26.4	10.0	4365	6	PCT-US91-08525-25	Sequence 25, Appl
34	26.4	10.0	4365	6	PCT-US93-04384-7	Sequence 7, Appl
35	26.4	10.0	4365	6	PCT-US93-04384-15	Sequence 15, Appl
36	26.4	10.0	4598	1	US-07-807-043B-5	Sequence 5, Appl
37	26.4	10.0	4598	1	US-08-299-849B-5	Sequence 5, Appl
38	26.4	10.0	4598	4	US-08-142-368A-5	Sequence 5, Appl
39	26.4	10.0	4598	5	US-08-967-727-5	Sequence 5, Appl
40	26.4	10.0	4757	1	US-07-865-662F-12	Sequence 12, Appl
41	26.4	10.0	6090	3	US-08-566-398-35	Sequence 35, Appl
42	26.4	10.0	6144	3	US-08-566-398-35	Sequence 35, Appl
43	26.4	10.0	35100	2	US-08-306-691B-19	Sequence 19, Appl
44	26.4	10.0	35100	6	PCT-US93-06251-19	Sequence 19, Appl
45	26.2	9.9	1512	5	US-09-142-514-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHETTLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PT.F15
US-08-232-463-14


```

? FILING DATE: 13-OCT-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: SANDERCOCK, COLIN G.
? REGISTRATION NUMBER: 31,298
? REFERENCE/DOCKET NUMBER: 18748/175
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202)672-5300
? TELEFAX: (202)672-5399
?
? TELEEX: 904136
? INFORMATION FOR SEQ ID NO: 33:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2575 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
?
? US-08-135-511-33

```

Query Match	10.9%	Score 28.8;	DB 1;	length 2575;
Best Local Similarity	52.5%;	Pred. NO. 1.2;		
Matches	63;	Conservative	0;	Mismatches 57;
			Indels	0;
			Gaps	0;

OY	124	tcgcctttaacaaacaaacacitacgccttcctcgatcaacattacgtgcctc	183
Db	1010	TCACATCAGAGAAAAAGCGCACATCTTCTCTCTCCCTCTTCTCTCCCTC	1066
OY	184	ttaaacgccttaagatcaaatgctcccaacagatgagaaatggaadtcctgaaga	243
Db	1070	TGACTGCCCTCTCTCATCTCTCTCTCTCTCCCTGACACTGCAAGGTAAATGTTGGCAGA	1122

RESULT 6
 US-08-483-852-10
 : Sequence 10, Application US/08483852
 : Patent No. 5650286
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: CHIANG, John Young Ling
 : TITLE OF INVENTION: Genomic DNA of Human Cholesterol
 : TITLE OF INVENTION: 7 α -Hydroxylase and Methods of Using It
 : NUMBER OF SEQUENCES: 11
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 :
 : COUNTRY: USA
 : ZIP: 20007-5109
 :
 : COMPUTER READABLE FORM:
 :
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION NUMBER: US/08/483,852
 : APPLICATION NUMBER: 07-JUN-1995
 : FILING DATE: 07-JUN-1995
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/361,458
 : FILING DATE: 21-DEC-1994
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/135,511
 : FILING DATE: 13-OCT-1993
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/135,488
 : FILING DATE: 13-OCT-1993
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/135,510
 : FILING DATE: 13-OCT-1993
 :
 : ATTORNEY/AGENT INFORMATION:
 : NAME: SANDERCOCK, Cold G.
 : REGISTRATION NUMBER: 31,298
 :
 : REFERENCE/DOCKET NUMBER: 18748/220 HOCE

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-852-10

Query Match 10.9%; Score 28.8; DB 1; Length 2575;
Best Local Similarity 52.5%; Pred. No. 1.2;
Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 124 tgcgtttaatcacaaactaacctatccgctccctccgcatcattactgcctc 183
DB 1010 TCACATTTCAGGAAAAAGCTCCACATATCTTCTCTCCCTCTCTCCCTCCCTCC 1069
QY 184 ttaaacctcccttaagatcaatgtcccaacagatgggaatggaagtcctgcaaga 243
DB 1070 TGACTGCCCTCTCTATCTCTCTCTCTCTCCCTGAGCTGGCAAGTTAATGTGCGACA 1129

RESULT 7
US-08-361-458-5
Sequence 5, Application US/08361458
Patent No. 5663483
GENERAL INFORMATION:
APPLICANT: Chiang, John
TITLE OF INVENTION: Genomic DNA of Human Cholesterol
TITLE OF INVENTION: 7a-Hydroxylase and Methods for Using it
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/361,458
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,488
FILING DATE: 13-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-361-458-5

Query Match 10.9%; Score 28.8; DB 1; Length 2575;

Best Local Similarity 52.5%; Pred. No. 1.2;
Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 124 tgcgtttaatcacaaactaacctatccgctccctccgcatcattactgcctc 183
DB 1010 TCACATTTCAGGAAAAAGCTCCACATATCTTCTCTCCCTCTCTCCCTCCCTCC 1069
QY 184 ttaaacctcccttaagatcaatgtcccaacagatgggaatggaagtcctgcaaga 243
DB 1070 TGACTGCCCTCTCTATCTCTCTCTCTCTCCCTGAGCTGGCAAGTTAATGTGCGACA 1129

RESULT 8
US-08-477-953-10
Sequence 10, Application US/08477953
Patent No. 5677159
GENERAL INFORMATION:
APPLICANT: CHIANG, John Young Ling
TITLE OF INVENTION: Genomic DNA of Human Cholesterol
TITLE OF INVENTION: 7a-Hydroxylase and Methods of Using It
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,953
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/361,458
FILING DATE: 21-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,511
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,488
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,510
FILING DATE: 13-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/219 HOCE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-953-10

Query Match 10.9%; Score 28.8; DB 1; Length 2575;
Best Local Similarity 52.5%; Pred. No. 1.2;
Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 124 tgcgtttaatcacaaactaacctatccgctccctccgcatcattactgcctc 183
DB 1010 TCACATTTCAGGAAAAAGCTCCACATATCTTCTCTCCCTCTCTCCCTCCCTCC 1069

OY 184 ttaaacctcctacagatcaaatgctcccaacagatgggaatggaatcctgcaaga 243
 Db 1070 TGACGTGCGCTCTTCTCTATCTCTCTCTCTCCCTGAGCTGGCAAGTTAATTGGTCGAGA 1129

RESULT 9 US-08-187-453-33

; Sequence 33, Application US/08187453
 ; Patent No. 5753431

GENERAL INFORMATION:

APPLICANT: Chiang, John

TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene

NUMBER OF SEQUENCES: 37
 REGULATORY ELEMENTS AND TRANSCRIPTION FACTORS

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/187,453

FILING DATE: 28-JAN-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/135,488

FILING DATE: 13-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/135,511

FILING DATE: 13-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/135,510

FILING DATE: 13-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: SANDERCOCK, COLIN G.

REGISTRATION NUMBER: 31,298

REFERENCE/DOCKET NUMBER: 18748/188

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ. ID NO. 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 2575 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-187-453-33

Query Match

Best Local Similarity 10.9%; Score 28.8; DB 2; Length 2575;
 Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 124 tcgctttatcaacaatacctatccgcttcctccgcatcattcaactgcctc 183
 Db 1010 TCACATTCAGGAAAAAGCTCCACTATCTTCTCTCTCCCTCTTCTCTCCCTCC 1069

OY 184 ttaaacctcctacagatcaaatgctcccaacagatgggaatggaatcctgcaaga 243
 Db 1070 TGACGTGCGCTCTTCTCTATCTCTCTCTCTCCCTGAGCTGGCAAGTTAATTGGTCGAGA 1129

RESULT 10

US-08-477-952-10

; Sequence 10, Application US/08477952

; Patent No. 5851780

GENERAL INFORMATION:

APPLICANT: CHIANG, John Young Ling

TITLE OF INVENTION: Genomic DNA of Human Cholesterol

NUMBER OF SEQUENCES: 11
 REGULATORY ELEMENTS AND METHODS OF USING IT

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,952

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/361,458

FILING DATE: 21-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/135,511

FILING DATE: 13-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/135,488

FILING DATE: 13-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/135,510

FILING DATE: 13-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: SANDERCOCK, COLIN G.

REGISTRATION NUMBER: 31,298

REFERENCE/DOCKET NUMBER: 18748/221 HOCE

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ. ID NO. 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2575 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-477-952-10

Query Match

Best Local Similarity 10.9%; Score 28.8; DB 3; Length 2575;
 Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 124 tcgctttatcaacaatacctatccgcttcctccgcatcattcaactgcctc 183
 Db 1010 TCACATTCAGGAAAAAGCTCCACTATCTTCTCTCTCCCTCTTCTCTCCCTCC 1069

OY 184 ttaaacctcctacagatcaaatgctcccaacagatgggaatggaatcctgcaaga 243
 Db 1070 TGACGTGCGCTCTTCTCTATCTCTCTCTCTCCCTGAGCTGGCAAGTTAATTGGTCGAGA 1129

RESULT 11

US-09-035-648-23

; Sequence 23, Application US/09035648

; Patent No. 6100031

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
 GROWTH AND PROLIFERATION

NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Fish & Richardson, P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: US
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/035,648
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/818,829
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Weiklejohn, Ph.D., Anita L.
 REGISTRATION NUMBER: 35,283
 TELEPHONE/DOCKET NUMBER: 07334/003001
 TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3131 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 449...2665
 US-09-035-648-23

Query Match 10.5%; Score 27.6; DB 5; Length 3131;
 Best Local Similarity 55.1%; Pred. No. 3.6;
 Matches 54; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 21 atggtcttcggttcacagctctcaagcgaatcacaattgcttccctatcagaagcgctt 80
 DB 1682 AAGTCCGTCATCTCCCAACAGTAACCTGTTCCCACTGCTGCTCCAGAGTGGCT 1741
 QY 81 cgcgcccgagattctgcgtcttcctccacttaaaccttc 118
 DB 1742 CCCACCAACCTCCTGCTTCCCACTCCACATCCACCCATC 1779

RESULT 12
 5340934-12/c
 Patent No. 5340934
 APPLICANT: TERMINE, JOHN D.; YOUNG, MARIAN F.; FISHER, LARRY W.
 ROBEY, PAMELA G.
 TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
 NUMBER OF SEQUENCES: 13
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/432,044
 FILING DATE: 03-NOV-1989
 SEQ ID NO: 12:
 LENGTH: 933
 5340934-12

Query Match 10.2%; Score 27; DB 7; Length 933;
 Best Local Similarity 53.3%; Pred. No. 3.2;
 Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 25 ctccggttcacagctctcaagcgaatcacaattgcttccctatcagaagcgcttcgc 84
 DB 211 CTTCGGTTCTCTCGACACATCATCAATTTCTTACTTCACACCTGGACAGATTAGCTC 152
 QY 85 ccgagattctgcgtcttcctccactcaaaccttcttcacatgccttt 131
 DB 151 CCACAGTACTCAGTCACTCCTCGCACAGTTTCTTCACACACTCT 105

RESULT 13
 US-08-821-355A-2
 Sequence 2, Application US/08821355A
 Patent No. 5851775
 GENERAL INFORMATION:
 APPLICANT: Barker, Nick
 APPLICANT: Clevers, Hans
 APPLICANT: Korinek, Vladimir
 APPLICANT: Morin, Patrice
 APPLICANT: Kinzler, Kenneth
 APPLICANT: Vogelstein, Bert
 APPLICANT: Sparks, Andrew
 TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
 TITLE OF INVENTION: Interact to Prevent Cancer
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, N.W.
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 20001
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/821,355A
 FILING DATE: 20-MAR-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A
 REGISTRATION NUMBER: 32,145
 TELEPHONE/DOCKET NUMBER: 1107,05064
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 TELEX: 97430 BMB UT
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2444 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-821-355A-2

Query Match 10.2%; Score 27; DB 3; Length 2444;
 Best Local Similarity 62.7%; Pred. No. 5.2;
 Matches 42; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 154 ctccctccgacattcactcgtccctcttaaacctccttaacgatacaatgctccc 213
 DB 107 CTCGCCCCCTTCCTCCCTCCTCTCTTCCCTCCCGACAGAAAGAACCCCC 166
 QY 214 aaacaga 220
 DB 167 AAGCAGA 173


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue Of The Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/518,878B
FILING DATE: 23-AUG-1995
CLASSIFICATION: 435

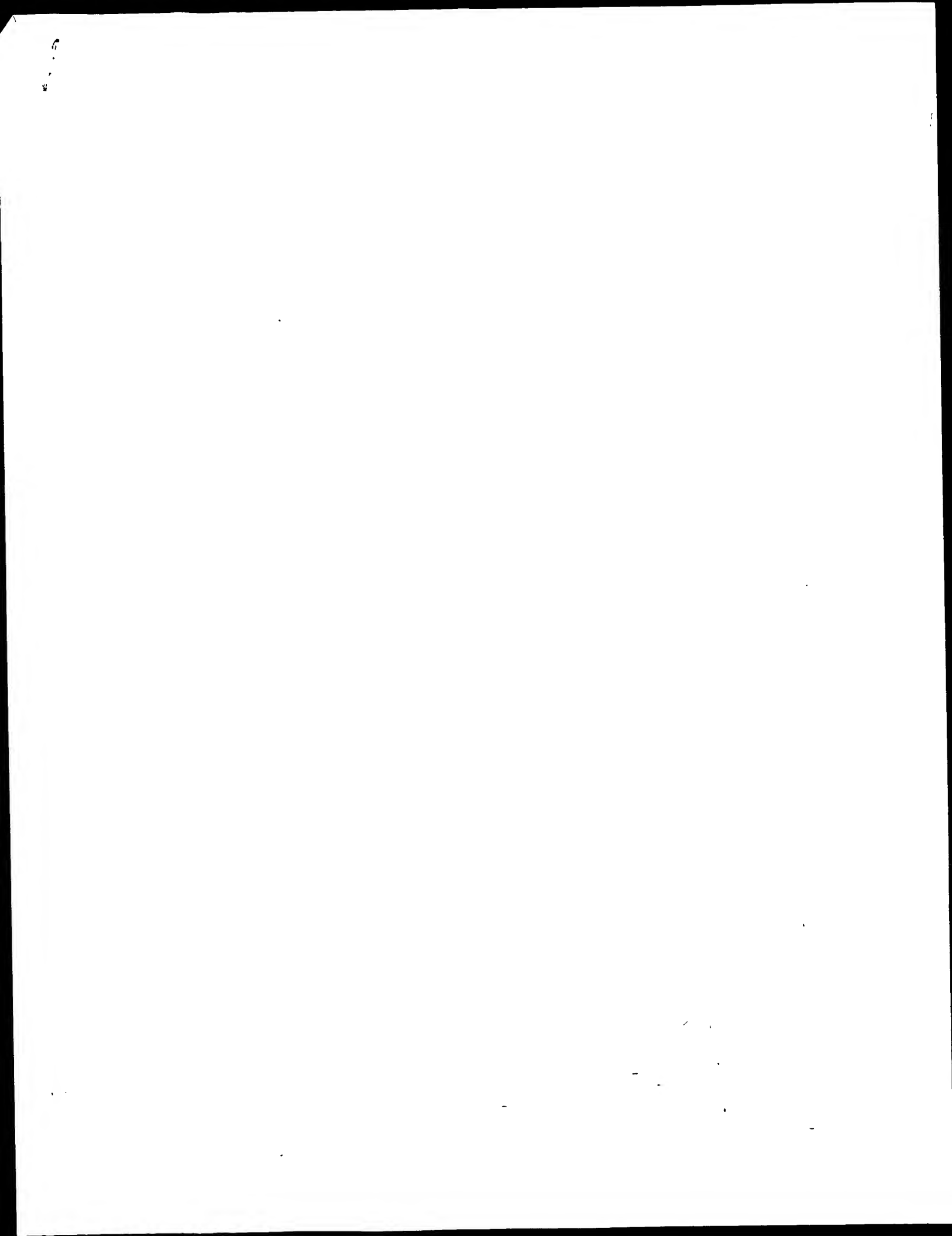
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-036
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

INFORMATION FOR SEO ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-518-878B-38

```

[illegible]

Search completed: November 11, 2000, 18:49:35
Job time: 2100 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:13:15 ; Search time 823.97 Seconds
(without alignments)
1980.971 Million cell updates/sec

Title: US-09-300-482-569
Perfect score: 264
Sequence: 1 ctgactgcagcaatcccca.....gacatgcttcgatcttta 264

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query Score	Match Length	DB	ID
---------------	----------------	--------------	----	----

1	252	95.5	540	20	AW234454	AW234454	s125d02.y
2	250.4	94.8	547	23	AW569249	s164g01.y	AW569249
3	250.4	94.8	565	23	AW569131	s163g01.y	AW569131
4	242	91.7	510	21	AW3995473	sh4fE02.y	AW3995473
5	230	87.1	435	10	AT1443659	s442h03.y	AT1443659
6	200.4	75.9	532	35	BE475226	pe73g04.y	BE475226
7	199.4	75.5	485	20	AW201037	se97b07.y	AW201037
8	199.4	75.5	485	20	AW201037	se97b07.y	AW201037
9	198.8	75.3	519	20	AW132188	sd67c11.y	AW132188
10	196.6	74.5	529	12	AT1748238	s49fH08.y	AT1748238
11	185.4	70.2	363	10	AT141411	s459d05.y	AT141411
12	158	59.8	452	24	AW830435	sm26g09.y	AW830435
13	147.4	55.8	452	24	BE347314	sp37e10.y	BE347314
14	139.6	52.9	460	35	BE347314	sp37e10.y	BE347314
15	96.8	36.7	461	35	BE440344	sp44d01.y	BE440344
16	67.8	25.7	684	23	BE210460	sc46a04.y	BE210460
17	67	25.4	599	23	AW685059	EST314548	AW685059
18	66.2	25.1	649	24	AW693590	NEF24G08N	AW693590
19	66.2	25.1	652	24	AW693590	NEF67H12S	AW693590
20	66.2	25.1	659	24	AW690852	NF035G07S	AW690852
21	66.2	25.1	661	34	AW691724	NP043E11S	AW691724
22	66.2	25.1	657	23	BE325705	NFO55G04S	BE325705
23	63.8	22.5	581	5	AW686396	NP011F06L	AW686396
24	59.4	22.5	502	23	AA660292	00161.MR	AA660292
25	59.4	22.5	724	24	AW585763	EST317386	AW585763
26	58.4	22.0	289	21	AW774002	EST332988	AW774002
27	49.2	18.6	657	12	AW306463	se51a10.y	AW306463
28	47.6	18.0	589	12	AT1726874	BNLGH1678	AT1726874
29	47.6	18.0	610	24	AT1729961	BNLGH1581	AT1729961
30	47.6	18.0	686	20	AW728667	GA_Ea001	AW728667
31	47.6	18.0	694	23	AW109134	gate0002H	AW109134
32	47.6	18.0	893	24	BE052281	GA_Ea000	BE052281
33	47.6	18.0	893	24	AW731245	GA_Ea003	AW731245
34	47.6	18.0	898	24	AW727712	GA_Ea001	AW727712
35	41.2	15.6	698	23	AW666886	GA_Ea000	AW666886
36	38.2	14.5	444	35	CNS0008TE	AT060767	Drosophili1
37	37.8	14.5	859	121	BE343965	BE343955	EST409117
38	37.4	14.2	795	35	BE034885	AT055406	Drosophili1
39	37.2	14.1	939	121	CNS000NG	BE034885	ML05F12.M
40	37	14.0	560	8	AT038541	AT059440	Drosophili1
41	37	14.0	597	14	AL042185	AT038541	ox3805.S
42	36.6	13.9	442	35	BE353752	AL042185	DKF5P434F
43	36.4	13.8	464	111	AOB83913	BE353752	EST355095
44	36.4	13.8	608	122	CNS02250	AOB83913	HS_4646.E
45	36.4	13.8	1303	35	BE421678	AT189431	retrodocto
						BE421678	HM0012CG.

RESULT
AW234454

LOCUS	540 bp	mrna	EST	17-JUL-2000
DEFINITION	AM234454	Glycine max cDNA clone	GENOME SYSTEMS	CLONE ID:
VERSION	Gm-c1028.y1	Gm-c1028	similar to	TR:004894 004894 TRANSALDOLASE ;, mrna
ACCESSION	AM234454			
KEYWORDS	sequence.			
SOURCE	AM234454.1	GI:6566822		
ORGANISM	soybean.			
REFERENCE	Glycine max			
AUTHORS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.			
	1 (bases 1 to 540)			
	Shoemaker,R., Keim,P., Vocklin,L., Erpelding,J., Corryell,V., Khanna			
	A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,			
	Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers			
	,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk			
	,R., Ratter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann			
	,R., Waterson,R. and Wilson,R.			
	Public Soybean EST Project			
	Unpublished (1999)			
TITLE	Public Soybean EST Project			
JOURNAL	Contact: Shoemaker R/Public Soybean EST Project			
COMMENT	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: est@watson.wustl.edu			
	This clone is available through: Genome Systems, Inc. 4633 World			
	Parkway Circle St. Louis, Missouri 63134 For further information			
	call: (800) 430-0030 or (314) 427-3322 FAX:(888) 919-3324 or (314)			
	427-3324 or contact: clones@genomesystems.com or			
	info@genomesystems.com web site: www.genomesystems.com			
	Insert Length: 1733 Std Error: 0.00			
	High quality sequence stop: 420.			
FEATURES	Location/Qualifiers			
source	1..540			
	/organism="Glycine max"			
	/db_xref="taxon:3847"			
	/clone="GENOME SYSTEMS CLONE ID: Gm-c1028-988"			
	/clone_1lb="Gm-c1028"			
	/tissue_type="roots of 'Supernod' plants"			
	/lab_host="DH10B"			
	/note="Vector: plunescrypt II XR; Site_1: EcoRI; Site_2:			
	XhoI; The mRNA was isolated from roots of Glycine max.			
	'Supernod' plants generously donated by Dr. Gary Stratey.			
	The seedlings were inoculated with Bradyrhizobium			
	japonicus, strain USDA110 prior to harvest. Stratey's			
	cDNA synthesis kit (catalog number 200401) was used to			
	synthesize the cDNA. First-strand synthesis was performed			
	with 5-methyl dCTP, hence the ligated cDNA was			
	hemimethylated. A modification of Stratey's			
	first-strand synthesis primer was used. An 'anchor'			
	nucleotide (V-A-C, or G) was added to the 3' end of the			
	primer (GAGAGACAGACAGACAGACAGACGTCTGAG(TT)18V) to anchor			
	the primer at the 5' end of the poly(A) tract. After			
	second-strand synthesis, the cDNA ends were filled in with			
	cloned Pfu DNA polymerase, ligated to EcoRI adaptors and			
	subsequently phosphorylated. The XhoI site within the			
	first-strand synthesis primer was then restricted by			
	digestion with XhoI; all XhoI sites in the cDNA would be			
	protected by their hemimethylated status. The cDNA			
	constructs were size-fractionated with a 500bp cutoff.			
	using GldcoBRL Lite Technologies' cDNA Size Fractionation			
	column. The column eluent was then ligated into			
	Stratey's plunescrypt II XR predigested vector			
	(DBlnescrypt II SR(+)) that has been digested with EcoRI			

ORIGIN

Query Match 95.5%; Score 252; DB 20; Length 540;
 Best Local Similarity 99.6%; Pred. No. 9.3e-68;
 Matches 263; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

2 tgacttgagcaatcccaatgcttcggttcgaagctcgaagcgaacatccactgc 61
 3 TGACTTGAGCAATCCCAATGCGCTTCGTTCCAGGCTCCCAAGCCCAATCCACTTGC 62
 62 ttccctatcagaagcgttcgccccgagatctcgtctccatccactcactcttc 121
 63 TTCCTTATCAGAAAGCGGTTGCGCCCGAGATTCTCGCTCCCTCCACTTCAACCTTCTTC 122
 122 catgctttaaatacaaaaactaatcctcgtctccctccatccactcactgc 181
 123 CATGCTTTTATATCAAAACTAATCATCGCTCCCTCCGATCCATCATCGCTGCC 182
 182 tcttaaaacctccttaagatcaatgctcccaacagatgggaatggaatctcga 241
 183 TCTTAAACCTCTTACGATCAATGCTCCCAACAGATGGGATGGAGATCGCTGCAA 242
 242 gaggaca-tgcttcatagtatctta 264
 243 GAGGACAGTGTCTCATGATCTTTA 266

RESULT 2

AM569249 547 bp mRNA EST 18-JUL-2000
 LOCUS AM569249
 DEFINITION s164901.y1 Gm-r1030 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-r1030-3721 5' similar to TR:004894 004894 TRANSALDOLASE ;, mRNA
 sequence.

ACCESSION AM569249
 VERSION AM569249.1 GI:7233907
 KEYWORDS EST.

SOURCE soybean.
 ORGANISM Glycine max

REFERENCE Eukaryota: Viridiplantae: Embryophyta: Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 547)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvett, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kuabara, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
 Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk,
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.

TITLE

JOURNAL

COMMENT

Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

FEATURES

source
 1..547
 Location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-r1030-3721"
 /clone_id="Gm-r1030"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; site_1: SalI; site_2: NotI; This
 High quality sequence stop: 438.
 infoGenomesystems.com web site: www.genomesystems.com
 427-3324 or contact: clones@genomesystems.com or
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 insert Length: 1708 Std Error: 0.00

CDNA library was constructed from mRNA isolated from
 immature cotyledons of greenhouse grown plants
 (individual seed fresh weight of 100-300mg). The library
 was prepared using the Life Technologies pSPORT1 cDNA
 library construction kit. Complementary DNA was
 synthesized from mRNA using a poly(dT) sequence with a
 NotI restriction site. SalI linker adapters were ligated
 to the blunt-ended cDNA fragments followed by NotI
 digestion. The cDNA fragments were directionally cloned
 into the NotI-SalI restriction site of the pSPORT1
 vector. The ligated cDNA fragments were transformed into
 E. coli Electromax DH10B host cells. This library was
 constructed by Dr. Lila Vodkin and Dr. Ann Khanna. Note
 that Gm-r1030 is a re-track of Gm-cl007."

Query Match 94.8%; Score 250.4; DB 23; Length 547;
 Best Local Similarity 99.2%; Pred. No. 2.9e-67;
 Matches 262; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

2 tgacttgagcaatcccaatgcttcggttcgaagctcgaagcgaacatccactgc 61
 10 TGACTTGAGCAATCCCAATGCGCTTCGTTCCAGGCTCCCAAGCCCAATCCACTTGC 69
 62 ttccctatcagaagcgttcgccccgagatctcgtctccatccactcactcttc 121
 70 TTCCTTATCAGAAAGCGGTTGCGCCCGAGATTCTCGCTCCCTCCACTTCAACCTTCTTC 129
 122 catgctttaaatacaaaaactaatcctcgtctccctccatccactcactgc 181
 130 CATGCTTTTATATCAAAACTAATCATCGCTCCCTCCGATCCATCATCGCTGCC 189
 182 tcttaaaacctccttaagatcaatgctcccaacagatgggaatggaatctcga 241
 190 TCTTAAACCTCTTACGATCAATGCTCCCAACAGATGGGATGGAGATCGCTGCAA 249
 242 gaggaca-tgcttcatagtatctta 264
 250 GAGGACAGTGTCTCATGATCTTTA 273

RESULT 3

AM569131 565 bp mRNA EST 18-JUL-2000
 LOCUS AM569131
 DEFINITION s163901.y1 Gm-r1030 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-r1030-3625 5' similar to TR:004894 004894 TRANSALDOLASE ;, mRNA
 sequence.

ACCESSION AM569131
 VERSION AM569131.1 GI:7233788
 KEYWORDS EST.

SOURCE

ORGANISM

Glycine max
 Eukaryota: Viridiplantae: Embryophyta: Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 565)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvett, V., Khanna,
 A., Bolla, B., Marra, M., Hillier, L., Kuabara, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
 Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk,
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.

TITLE

JOURNAL

COMMENT

Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1744 Std Error: 0.00
High quality sequence stop: 415.

FEATURES

source

1. 565
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-r1030-3625"
/clone_id="Gm-r1030"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site: 1: SalI; Site: 2: NotI. This cDNA library was constructed from mRNA isolated from immature cotyledons of greenhouse grown plants (individual seed fresh weight of 100-300mg). The library was prepared using the Life Technologies pSPORT1 cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linkers followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Ann Khanna. Note that Gm-r1030 is a re-track of Gm-cl007."

BASE COUNT 154 a 145 c 112 g 153 t 1 others
ORIGIN

Query Match 94.8%; Score 250.4; DB 23; Length 565;
Best Local Similarity 99.2%; Pred. No. 2.9e-67;
Matches 262; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 tgaattgagcaatcccaatgcttccttcgaagctcgaagccaatccacttgc 61
DB 10 TGACTTCACCAATTCCTCCGTTTAAAGCTCTCAACGCCAATCCACTTGC 69
QY 62 ttcttcaagaagcgttcgccccgagattcgtcttcctcaactcaaacctcttc 121
DB 70 TTCTTTTCAGAAAGCGTTTGGCCCGAGATTCCTCGTTCACTTCAACCTTCTTC 129
QY 122 catgctttaatacaaaactaactatcgcgttccttcgaatcaatcgtctcc 181
DB 130 CATGCTTTTAATCAAAACTAATCATTCGCTTCCTCGATTCATTCACCTGCTCC 189
QY 182 tcttaaaactccttaagataatgctccaaacagatgggaatggaatcgtcaaa 241
DB 190 TCTTAAACCTCTTACGATCAATGCTCCCAACAGATGGATGGAAAGTCTCGAAA 249
QY 242 gaagaca-tgctcatgatcttta 264
DB 250 GAGGACAGTGTCTCATGATCTTTA 273

RESULT 4
AM395473 510 bp mRNA EST 17-JUL-2000
LOCUS sn49f02.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl017-5044 5' similar to TR:004894 004894 TRANSALDOLASE ;, mRNA
sequence.

ACCESSION AM395473
VERSION AM395473.1 GI:6913943
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 510)
Eukaryota, Viridiplantae, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, kosiidae, eurosids I, Fabales, Fabaceae, Papilionoideae, Glycine.

AUTHORS

Shoemaker R., Keim P., Vodkin L., Erpelting J., Corryell V., Khanna A., Bolla B., Marra M., Hillier L., Kueba T., Martin J., Beck C., Wylie T., Underwood R., Steptoe M., Theising B., Allen M., Bowers Y., Person B., Swaller T., Gibbons M., Page D., Harvey N., Schirck R., Ritter E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann R., Waterston R. and Wilson R.
Public Soybean EST Project
Unpublished (1999)

JOURNAL

Contact: Shoemaker R./Public Soybean EST Project

COMMENT

Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1816 Std Error: 0.00
Seq primer: -40RP from Glibco
High quality sequence stop: 395.

FEATURES

source

1. 510
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-5044"
/clone_id="Gm-cl017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: Bluescript II XR; Site: 1: EcoRI; Site: 2: XhoI. This cDNA library was constructed from mRNA isolated from vegetable buds of field grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 141 a 138 c 98 g 133 t
ORIGIN

Query Match 91.7%; Score 242; DB 21; Length 510;
Best Local Similarity 99.6%; Pred. No. 1.2e-64;
Matches 253; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 12 caattcccaatgcttcgcttcgaagcttcaagcaaaatccactgcttcctatca 71
DB 1 CAATTCCCAATGCTTCGTTTCAAGCTCTCAAGCCCAATCCACTGCTTCCTATCA 60
QY 72 gaagcgttcgccccgagattcgtcttcctcaacttcaaaccttctcatcgcttt 131
DB 61 GAAGCGCTTCGCCCCGAGATTCCTGCTTCCTCACTTCAAACTTCTTCCATGCTTTT 120
QY 132 aatcaaaaactaactatcgcgttccttcgaatcattcactgcttcctcttaaac 191
DB 121 AATCACAAAACTAATCACTATCCGCTTCCTCCGATCCATTCACCTCCCTTTAAAC 180
QY 192 tcttaagataaattgctcccaaaagatgggaatggaatcgtcaaaagagaca-tg 250
DB 181 TCTTACGATCAAAATGCTCCCAACAGATGGATGGAAAGTCTCGCAAAAGGACACTG 240
QY 251 ctccatgatcttta 264
DB 241 CTTCATGATCTTTA 254

RESULT 5
 A1443659 435 bp mRNA EST 18-APR-2000
 LOCUS
 DEFINITION sa42h03.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl004-1998 5' similar to TR:004894 004894 TRANSALDOYLASE ;, mRNA
 sequence.
 A1443659
 A1443659.1 GI:4304718
 EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 435)
 REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Rhanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., Mccann
 ,R., Waterson,R. and Wilson,R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Possible reversed clone: similarity on wrong strand
 Insert Length: 1680 Std Error: 0.00
 Seq primer: -40RP from GIBCO
 High quality sequence stop: 353
 POLYA=NO.

FEATURES

source
 1.435
 Location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl004-1998"
 /clone_1ib="Gm-cl004"
 /tissue_type="root"
 /lab_host="XLI0-GOLD"
 /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; Root cDNA. The mRNA was isolated from entire roots
 of 8 day old 'Williams' seedlings which were propagated on
 paper towels with distilled water. StrataGene's cDNA
 Synthesis Kit (catalog #200401) was used to synthesize the
 cDNA. First- strand synthesis was performed with 5-methyl
 dCTP, hence the ligated cDNA is hemimethylated.
 StrataGene's first-strand synthesis was used
 [GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG(T)-18]. After
 second-strand synthesis, the cDNA ends were 'polished',
 with clone pfu DNA polymerase, ligated to EcoRI adapters,
 and phosphorylated. The XhoI site within the first-strand
 synthesis primer was restricted by digestion with XhoI;
 all XhoI sites in the cDNA would be protected by their
 hemimethylated status. The cDNA constructs were
 size-fractionated with a 500bp cutoff, using GIBCOBRL life
 technologies' cDNA Size Fractionation column. The column
 eluent was then ligated into StrataGene's pBluescript II
 XR Predigested vector (pBluescript II SK(+)) that had been
 digested with EcoRI and XhoI, and phosphorylated. Both
 the white and blue colonies appear to contain recombinant
 plasmids with cDNA inserts. Blue colonies 9n-15) have been
 sequenced, and possess putative cDNA inserts. This library
 was constructed by Dr. Paul Keim & Virginia H. Coryell,

Department of Biology, Box5640, Northern Arizona
 University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
 Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
 520-523-7500, email: paul.keim@nau.edu,
 virginia.coryell@nau.edu
 BASE COUNT 116 a 126 c 79 g 114 t
 ORIGIN

Query Match 87.1%; Score 230; DB 10; Length 435;
 Best local Similarity 99.6%; Pred. No. 5.8e-61;
 Matches 241; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 24 gcttcggttccaagctcacaagccaatcactgtcttcttcaagaagcgttcg 83
 DB 1 GCTTCGTTCCAAAGCTCACAAGCCAATCCACTGCTCTTATCAGACGCTTGC 60
 QY 84 ccccgagattctgctctctcaactcaaccttcttcacgtcttcaatcaaaaact 143
 DB 61 CCCGAGATTCTGCTCTCCTCACTCAAAACCTTTCATCGTTTAATCAAAACT 120
 QY 144 aaactatccgcttccctccgacatccatcactgcctcccttaaaactccttaagatc 203
 DB 121 AACCTATCGGCTTCCTCCGATCCATTCCTGCTCCCTTAAACCTCTTACGATC 180
 QY 204 aaatgcctcccaacagatggaatggaatgctctgcaagaagagaca-tgctcatatctt 262
 DB 181 AATGCTCCCAAAACAGATGGAATGGAAGTCTCGAAGAGACAGAGCTCTCATCTT 240
 QY 263 ta 264
 DB 241 TA 242

RESULT 6
 BE475226 532 bp mRNA EST 28-JUL-2000
 LOCUS
 DEFINITION sp73g04.y1 Gm-cl044 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl044-1615 5' similar to TR:004894 004894 TRANSALDOYLASE ;, mRNA
 sequence.
 BE475226
 BE475226.1 GI:9565717
 EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 532)
 REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Rhanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., Mccann
 ,R., Waterson,R. and Wilson,R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 360.
 location/Qualifiers
 1.352
 /organism="Glycine max"

FEATURES

source

/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1044-1615"
/clone_id="Gm-c1044"
/tissue_type="Hypocotyl, 9-10 day old etiolated seedlings"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+, Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from etiolated hypocotyl tissue of 9-10 day old seedlings of the cultivar Williams. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) primer with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestion with EcoRI and XhoI. The cDNA-XhoI fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

BASE COUNT 144 a 137 c 105 g 146 t

ORIGIN

Query Match 75.9%; Score 200.4; DB 35; Length 532;
Best Local Similarity 92.9%; Pred. No. 9, 3e-52;
Matches 221; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 28 ccgttcccaagctcgaacgaacatccactgtcttccttaccagaagcgtcgcgcc 87
DB 6 CCATTTCAGAGCTCAACGCCATCTCCAGCTCTTATCAGAAAGCGCTTCGCCCA 65
QY 88 gaagatcgtcttcctcactcctcaaaccttccttcacttcaacaaactacc 147
DB 66 GAGATTCTCCCTCTCTCTTCAATCTCTTCCACAGCTATTATTAACAAATTAAC 125
QY 148 tatccgtcttcctcgaatcattcactcgtcctcttaaaccttcactgaatcaat 207
DB 126 TCTCCGCTTCCCTCCATTCATTCACCTCCCTTTAAACCTTCGGGATCAAT 185
QY 208 gctcccaacagatgggaatggaagtcctcgaagagaga-tgcttcattgacttta 264
DB 186 GCTCCCAAGAGATGGAATGGAATGCTCCCAAGAGAGACAGTTCATGATCTTTA 243

RESULT 7

AM201037

LOCUS

DEFINITION

690 bp mRNA EST 17-JUL-2000
se97b07.y1 Gm-c1027 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1027-686 5' similar to YR:004894 004894 TRANSALDOLASE ; mRNA sequence.

ACCESSION

AM201037

KEYWORDS

SOURCE

ORGANISM

soybean.

glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE

1 (bases 1 to 690)

Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shuh, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

PUBLIC SOYBEAN EST PROJECT

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1881 Std Error: 0.00
High quality sequence stop: 422.
Location/Qualifiers

FEATURES

source

1. 690
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1027-686"
/clone_id="Gm-c1027"
/tissue_type="cotyledons of 3- and 7-day-old Williams seedlings"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+, Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 3- and 7-day-old Williams seedlings which were propagated on paper towels with distilled water. The cotyledons were flash-frozen in liquid nitrogen, then lyophilized for 72 hours. Unequal amounts of mRNA was used for cDNA synthesis. StrataGene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of StrataGene's first-strand synthesis primer was used. An anchor nucleotide (V-A-C, or G) was added to the 3' end of the primer (GAGAGAGAGAGAGAGAGAGACTATCTCGAGT) to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated on a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into StrataGene's pBluescript(II) M13 digested vector (pBluescript II SK(+)) that has been digested with EcoRI and XhoI, and phosphorylated by StrataGene. 97% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=30). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 189 a 161 c 145 g 192 t 3 others

ORIGIN

Query Match 75.5%; Score 199.4; DB 20; Length 690;
Best Local Similarity 92.8%; Pred. No. 2e-51;
Matches 220; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 29 ccgttcccaagctcgaacgaacatccactgtcttccttaccagaagcgtcgcgcc 88
DB 6 CCATTTCAGAGCTCAACGCCATCTCCAGCTCTTATCAGAAAGCGCTTCGCCCA 65
QY 89 agatcgtcttcctcactcctcaaaccttccttcactgcttaccacaaactacc 148
DB 66 GAGATTCTCCCTCTCTCTTCAATCTCTTCCACAGCTATTATTAACAAATTAAC 125
QY 149 atccgttccctcgaatcattcactcgtcctcttcaaaccttcactgaagataatg 208
DB 126 CTCGCTTCCCTCCATTCATTCACCTCCCTTTAAACCTTCGGGATCAATG 185
QY 209 cttcccaacagatgggaatggaagtcctcgaagagaga-tgcttcattgacttta 264
DB 186 CTCCTCAAGAGATGGAATGGAATGCTCCCAAGAGAGACAGTTCATGATCTTTA 242

RESULT 8

BE211683
LOCUS 485 bp mRNA EST 29-JUN-2000
DEFINITION s064h11.y1 Gm-c1040 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1040-478 5' similar to TR:004894 004894 TRANSALDOLASE ; mRNA sequence.
ACCESSION BE211683
VERSION BE211683.1 GI:8827953
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 485)
Shoemaker, R., Keim, P., Vodkin, L., Erpellding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 280.
FEATURES
source
location/Qualifiers
1..485
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1040-478"
/clone_lib="Gm-c1040"
/tissue_type="Hypocotyl and Plumule, germinating seeds"
/lab_host="DH10B"
/note="Vector: p773Pac (Pharmacia); Site: 1: EcoRI; Site: 2: NotI; This cDNA library was constructed from mRNA isolated from hypocotyl and plumule tissues of seeds germinated for three days of the cultivar Williams 82. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a NotI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestions with EcoRI and NotI. The cDNA fragments were directionally cloned into the EcoRI-NotI restriction site of the p773-Pac vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

BASE COUNT 128 a 129 c 102 g 125 t 1 others
ORIGIN

Query Match 75.4%; Score 199; DB 34; Length 485;
Best Local Similarity 93.2%; Pred. No. 2.5e-51;
Matches 219; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 31 ttccaagctcacaacccaatccaactgtctctcattacagaagcgttcgccccgag 90
DB 1 ttttcaagctcacaacccaatccaactgtctctcattacagaagcgttcgccccgag 60
QY 91 attcgtctcctcaccctcaaaccttcttcacgtcttatacacaactaacctat 150
DB 61 atttcgtctcctcaccctcaaaccttcttcacgtcttatacacaactaacctat 120
QY 151 ccgcttcctcctcagatcattcactcgtctcctccttaaaaccctcctacgaatgct 210

DB 121 cccgcttcctcctcagatcattcactcgtctcctccttaaaacctcttcgcatataatgct 180
QY 211 cccaacagatgggaatggagtcctgcgaagagaca-tgcttcattcattta 264
DB 181 cccaacagatgggaatggagtcctgcgaagagagacagtcgttcattcattta 235

RESULT 9
LOCUS AM132188
DEFINITION s067c11.y1 Gm-c1008 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1008-1389 5' similar to TR:004894 004894 TRANSALDOLASE ; mRNA sequence.
ACCESSION AM132188
VERSION AM132188.1 GI:6133795
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 519)
Shoemaker, R., Keim, P., Vodkin, L., Erpellding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gldco
High quality sequence stop: 449.
FEATURES
source
location/Qualifiers
1..519
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1008-1389"
/clone_lib="Gm-c1008"
/lab_host="DH10B"
/note="Vector: pSPOR1; Site: 1: SalI; Site: 2: NotI; This cDNA library was constructed from mRNA isolated from whole young pods, approximately 2cm long, of 12-week-old greenhouse grown plants. The library was prepared using the Life Technologies superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPOR1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

BASE COUNT 141 a 135 c 101 g 142 t
ORIGIN

Query Match 75.3%; Score 198.8; DB 20; Length 519;
Best Local Similarity 92.5%; Pred. No. 2.9e-51;

Query Match	70.2%	Score 185.4;	DB 10;	Length 363;
Best Local Similarity	92.0%	Pred. No. 3.8e-47;		
Matches 206;	Conservative 0;	Mismatches 17;	Indels 1;	Gaps 1;
QY 42	lcaacgccaactlccactgtctcttccatcagaagcgctgcgcccgagattctcgctc	101		
Db 1	TCAACGCCATCTCCAGCGCTCTTCCCTATCAACAACGCTTCCGCCAGAGATTCTGGCTC	60		
QY 102	ctcaactccaacctcttctccatgctttaaatacaaaaactatactcgcttcctc	161		
Db 61	CTCTCTTCAATCCCTTCTCCACGCGTATTAATTAACAANAATTAACCTCTCGCTTCCCTC	120		
QY 162	cgatccattctactcgcttcctccctcttaaacctcccttagcgatccaatgtcccaaacgat	221		
Db 121	CGATCCATTCTACTCCTCCCTCCCTCTTAANAACCTCTTGGGATCAAAATCTCCCAACAGAT	180		
QY 222	gggaagtgggaagtcctctgcgaagaggaca-tgcttaigtatcctta	264		
Db 181	GGGAATATGAAGTCTCTGCANAGAGGACAGTGCCTCATGTATCTTTA	224		
RESULT 12				
AM830435				
LOCUS	AM830435	452 bp	mRNA	
DEFINITION	sm2b6909.y1 Gm-c1028 glycine max cDNA clone GENOME SYSTEMS CLONE ID:		EST	18-JUL-2000
	Gm-c1028-5081 5' similar to YR:004894 004894 TRANSALDOLASE ;, mRNA			
ACCESSION	sequence-			
VERSION	AM830435			
KEYWORDS	AM830435.1 GI:7924409			
SOURCE	EST.			
ORGANISM	soybean.			
	Glycine max			
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;			

REFERENCE	1	Fabales; Fabaceae; Papilionoideae; Glycine.
AUTHORS		1 (bases 1 to 452)
TITLE		Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corceill, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shiu, T., Jackson, Y., Cardenas, M., Mccann, R., Waterson, R. and Wilson, R.
JOURNAL		Public Soybean EST Project
COMMENT		Unpublished (1999) Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu ↓ This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Insert length: 1169 Std Error: 0.00 High quality sequence stop: 386. Location/Qualifiers 1..452
FEATURES		
source		

BASE COUNT	134	a	111	c	90	g	117	t
ORIGIN								
Query Match	59.8%; Score 158; DB 24; Length 452;							
Best Local Similarity	99.4%; Pred. No. 1.2e-38;							
Matches 169; Conservative	0; Mismatches 0; Indels 1; Gaps 1;							
QY	96	cgcttcctacactcaaacctcttccatgcgtttaatacacaacctaccatccgct	155					
Db	6	CGCTTCGCACTTCAAACTTTTCCATGCCCTTTAATCAACAAAACTAACCTTATCCGT	65					

QY	156	tcocccagatcattcactgcctccctcttaaacctcccttaagatccaatgtccca	215
Db	66	TCCTCCGATTCATTCTACTGCTCCCTCTTTAAACCTCTTACGATCAATGTCCCA	125
QY	216	acagatgggaatggaagtcctgcgaagaagaca-tgcttaatgacttta	264
Db	126	ACAGATGGGAATGGAAGTCCTGCAAGAGACAGTCTTCATGATCTTTA	175
RESULT	13		
LOCUS	BE347314		
DEFINITION	690 bp mRNA	EST	18-JUL-2000
VERSION	sp37e10.y.1	Gm-cl043 glycine max CDNA clone	GENOME SYSTEMS CLONE ID:
KEYWORDS	Gm-cl043-451 5' similar to tr:004894 004894	TRANSLDOLASE	; mRNA
SOURCE	sequence.		
ORGANISM	BE347314		
REFERENCE	BE347314.1	GI:9259167	
AUTHORS	soybean.		
	Glycine max		
	Eukaryota, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;		
	Fabales; Fabaceae; Papilionoideae; Glycine.		
	1 (bases 1 to 690)		
	Shoemaker, R., Kelm, P., Vodkin, L., Erpelding, J., Coryell, V., Khana		
	, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,		
	Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers		
	, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk		
	, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann		
	, R., Materson, R. and Wilson, R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Shoemaker R/Public Soybean EST Project		
	Public Soybean EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: estewatson.wustl.edu		
	This clone is available through: Genome Systems, Inc. 4633 World		
	Parway Circle St. Louis, Missouri 63134 For further information		
	call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)		
	427-3324 or contact: clones@genomesystems.com or		
	info@genomesystems.com web site: www.genomesystems.com		
	High quality sequence stop: 485.		
FEATURES	Location/Qualifiers		
source	1..690		
	/organism="Glycine max"		
	/db_xref="taxon:3847"		
	/clone="GENOME SYSTEMS CLONE ID: Gm-cl043-451"		
	/clone_1lb="Gm-cl043"		
	/tissue_type="Hypocotyl and Plumule, germinating seeds"		
	/lab_host="DH10B"		
	/note="Vector: p773Pac (Pharmacia); Site:1: EcoRI;		
	Site:2: NotI; This CDNA library was constructed from mRNA		
	isolated from hypocotyl and plumule tissues of seeds		
	germinated for three days of the cultivar Williams.		
	Complementary DNA was synthesized from mRNA using a primer		
	consisting of a poly(dT) sequence with a NotI restriction		
	site. EcoRI adapters were ligated to the blunt-ended cDN		
	fragments followed by digestion with EcoRI and NotI. The		
	cDNA fragments were directionally cloned into the		
	EcoRI-NotI restriction site of the p773-Pac vector. The		
	ligated cDNA fragments were transformed into DH10B host		
	cells (Gibco BRL). This library was constructed by Dr.		
	Randy Shoemaker."		
BASE COUNT	185 a	168 c	149 g
ORIGIN	187 t	1 others	
Query Match	55.8%;	Score 147.4;	DB 35; Length 690;
Best Local Similarity	90.8%;	Pred. No. 2.7e-35;	
*Matches 157; Conservative	0;	Mismatches 16;	Indels 0; Gaps 0

QY	27	tcgattccagctcctaagcccaatccatctctctccatccagaagcgctccccc	86
Db	14	ttccattttcaaccttcctcaacgccatctccacgtctctcttatacagaaggcttccccc	73
QY	87	cgagattctcgcttccctacacctccaacacctcttcacatcgctcttataccacaacaaactaac	146
Db	74	AGAGATTCTCGCTTCTCTCTCTCTCAATCCCTCTTCCCAACGCTATTATATGCAAAATTATAC	133
QY	147	ctatccgcttccctccgatccatccatccatctgcgtccctccctttaaaccctccctacg	139
Db	134	CTCTCCGCTTCCCTCCGATTCATTCACATTCACATCCCTCCCTTAAACCTCTTGGC	186
RESULT	14		
LOCUS	BE440344		
DEFINITION	BE440344	461 bp	mRNA
ACCESSION	BE440344		
VERSION	BE440344.1	GI:9439831	
KEYWORDS	EST.		
ORGANISM	soybean.		
SOURCE	Glycine max		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
SOURCE			
BASE COUNT	134 a	104 c	94 g 129 t
ORIGIN			

were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

BASE COUNT	111 a	77 c	56 g	78 t
ORIGIN				

Query Match	36.7%;	Score	96.8;	DB	34;	Length	322;
Best Local Similarity	84.0%;	Pred	No	1	10-10;		

Matches	121;	Conservative	0;	Mismatches	22;	Indels	1;	Gaps	1;
---------	------	--------------	----	------------	-----	--------	----	------	----

QY 122 catgcgtttaatcaaaactaacctatccgcgttccctccgataccattcaatcgcctccc 18

Db
1 CAACACTATTAATTCGAATTAACCTCTCCGGTTCCTCCAAATCATTCACCTCCCTCC 60

QY 182 tcttaaacctccttaccgaatcaaatattcccaaagacatccaaataccaaatactggttt

Dbb

61 TCTTAAAAACCTCCCTCCTGATGCCAATTCAAAATTCAGGGAGCGGTTCCCTGCATAA 24

CCCTTGGGTTTCAGATGACCAATGCGAATTGAAAGAGGTGATCCTGCTGCATAA
12

242 gagggac-atgcttcgatcttta 26

Search completed: November 11, 2000, 18:47:51
Job time: 2076 sec

REFERENCE

AUTHORS

TITLE

**JOURNAL
COMMENT**

COMMENT

FEATURES

Source

1

